12-2015

# Subfunctionalization of Ethylene Receptors and Homology Modeling of Cytosolic Domains in *Arabidopsis thaliana*

Sai Keerthana Wuppalapati
*University of Tennessee - Knoxville*, swuppala@vols.utk.edu

To the Graduate Council:

I am submitting herewith a thesis written by Sai Keerthana Wuppalapati entitled "Subfunctionalization of Ethylene Receptors and Homology Modeling of Cytosolic Domains in *Arabidopsis thaliana*." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Biochemistry and Cellular and Molecular Biology.

Brad M. Binder, Major Professor

We have read this thesis and recommend its acceptance:

Barry Bruce, Jerome Baudry

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

# Subfunctionalization of Ethylene Receptors and Homology Modeling of Cytosolic Domains in *Arabidopsis thaliana*

**A Thesis Presented for the**
**Master of Science**
**Degree**
**The University of Tennessee, Knoxville**

**Sai Keerthana Wuppalapati**
**December 2015**

# Dedication

Dedicated to **MY FAMILY** and **Dr. BRAD M. BINDER**

# Acknowledgements

This project would not have been a reality without the Department of Biochemistry Cellular and Molecular Biology at the University of Tennessee, Knoxville, for giving me an opportunity to study here and for providing the brilliant laboratory equipments, computational and library facilities that helped me in developing a research oriented attitude. I appreciate the infrastructure of this institute which helps every student to rekindle their passion for research and for developing the right scientific attitude.

My heartfelt gratitude and deepest thanks to my advisor, **Dr. Brad M. Binder** for being a great mentor and a constant support of inspiration through the project. I am thankful to him for providing me with an enriching experience in all spheres during the project and being patient in correcting my flaws and giving this thesis a final form. I have benefited significantly from his experience and expertise in plant physiology and scientific writing. This would not have been possible if not for his encouragement.

I would like to thank our collaborator, my committee member **Dr. Jerome Baudry,** for his critical inputs that have been very important for the successful completion of this project. I have under his guidance learnt to strengthen my basics in computational structural biology, with an emphasis on molecular modeling.

My sincere and humble thanks to my committee member, **Dr. Barry D. Bruce** for having encouraged me at every juncture of this project, for having been so approachable to exchange any point of view and for instilling great hopes about the future in the field of science.

**Dr. Elizabeth Howell** needs a special mention for being my role model, for her encouragement expertise and advice and her snippets of knowledge that has contributed to my intellectual growth

I wish to express my gratitude and appreciation to Binder lab mates especially our plant fairy **Dr. Rebecca L. Wilson,** my mini mentor **Randy F. Lacey ,** my close friend **Arkadipta Bakshi** for their support and making my lab time the best time of the day. Binder lab will be missed very much. I would also like to thank **Karan Kapoor** from Baudry lab and **Khushboo Bafna** from Agarwal lab for their timely advice and critical inputs into the molecular modeling project.

My heartfelt thanks to my encouraging and optimistic friends **Punita Manga, Prashasti Kumar, Deepika Karunakaran, Jagadish Cherukuri, Dinesh Patlolla, Michelle Wieczynski, Rekha DeSilva, Sarvesh Iyer, Snehal Joshi, Sagar Utturkar, Sagar Yadavali, Siva Karthik Varanasi, Abhishek Chimalapati, Guru Venkatesan, Ansul Lokdarshi, Craig Conner, Willy Evangelista, Andrew Stafford, Tiffany Thoms, Tian Li, Emily Bean** who have grown with me for the past three years. I express my appreciation for their constructive criticisms during my course of study and their moral support. Thank you for having banked upon me through the tough times, assuring me in every stage and never letting my spirits sink.

I am deeply in debt and forever grateful to my father **Sri Murali Krishna Wuppalapati**, my mother **Smt. Chengalamma Kaduru,** my brother **Vamsi Krishna Wuppalapati,** my sister-in-law **Smt. Shilpa Madhuri Wuppalapati** and my nephew **Aniketh Krishna Wuppalapati** and other family members for their unconditional love. You constantly motivated me with undaunted support and tolerance that led me to conceive and complete this thesis in its full capacity and crafted out its productivity. Your selfless sacrifice, love and constant prayers are the reasons for completion of my dissertation.

With greatest reverence and gratitude I thank the **Almighty Lord** for bestowing me the gift of life and the precious opportunity to be a part of his divine mission. Lord, you have guided me through the phases of my life and you have never left my hand. You trusted me with the

capability to take up this project which had challenging protocols and you were there to give enormous amounts of strength to fight my failures. To you my dear lord, I dedicate my work and surrender myself to you.

# Abstract

Ethylene is a gaseous phytohormone that initiates and modulates several mechanisms related to growth and development in plants through a family of five disulphide-linked receptor dimers. Although the ethylene receptors are very similar in their structures, they have diverse functions with both overlapping and non-overlapping roles. Silver ions are able to support ethylene binding to the receptors but it is also interesting to note that ethylene responses are blocked in the presence of silver. A part of the present study identified that ETR1 receiver domain has little or no role in mediating responses to silver ions, supported by data obtained from end point analysis and analyzing growth kinetics of dark grown *Arabidopsis* seedlings. However, previous data suggested that these receptors are important for other responses. This led to an interest in studying the structural aspects that lead to the sub-functionalization of ethylene receptors. The current study mainly focuses on looking at the structures of these domains for a better understanding of their physiological roles. As information regarding the crystal structures of different domains of ethylene receptors is only limited to ETR1 catalytic domain and receiver domain, we predicted the three dimensional protein structure using knowledge-based prediction, homology modeling.

The models generated for receiver domains showed similar tertiary structure for ETR2 and EIN4 receiver domains as compared to that of ETR1 crystal structure. The models created for kinase domains suggested that although the sub families function through different kinases, structurally they were similar to sensor histidine kinases. ERS2 had been an exception for this and the DHp domain of ERS2 is yet to be characterized well. The models predicted for GAF domains suggested that GAF domains mostly have conserved alpha helices and the models generated either long loops or very short beta strands against beta strands in the templates. We could

predict the approximate structures of the different receptor domains and compared each of the predicted structures of each domain in all the receptor isoforms for a better understanding of how conformational changes in the structure result in different physiological functions.

# Table of Contents

1 INTRODUCTION ..... 1
1.1 Ethylene as a Phytohormone ..... 1
1.2 Ethylene Receptors in *Arabidopsis thaliana* ..... 2
1.3 Ethylene Signal Transduction Pathway ..... 2
1.4 Role of Metal Ions in Ethylene Receptor Function and Signaling ..... 3
1.5 Effect of Silver Nitrate on Ethylene Growth Responses of Dark-Grown *Arabidopsis* Seedlings ..... 3
1.6 Sub Functionalization of Ethylene Receptors ..... 4
1.7 Molecular Modeling Using Computational Approaches ..... 5
1.8 Homology Modeling of Proteins ..... 6
1.9 Goal of the Current Study ..... 6
2 MATERIALS AND METHODS ..... 7
2.1 Seedling Preparation ..... 7
2.2 Growth-Rate Measurements of Hypocotyls ..... 7
2.2.1 End Point Analysis ..... 7
2.2.2 High Resolution Time Lapse Imaging ..... 8
2.3 Homology Modeling of Ethylene Receptors ..... 8
2.3.1 Template Identification ..... 9
2.3.2 Structure Preparation ..... 9
2.3.3 Amino Acid Sequence Alignment and Correction ..... 9
2.3.4 Backbone Generation ..... 10
3 RECEIVER DOMAIN ..... 11
3.1 Introduction ..... 11
3.2 Role of ETR1 Receiver Domain in Responses to Silver ..... 12
3.2.1 Role of Receiver Domain in Ethylene Blocking Effects of Silver Nitrate ..... 12
3.3 Homology Modeling of Cytosolic Domain of Ethylene Receptors ..... 13
3.3.1 Sequence Alignment and Similarity Between Receiver Domains of ETR1, ETR2 and EIN4. ..... 14

3.3.2 Predicting Three Dimensional Structure of ETR2 Receiver Domain..... 15
3.3.3 Homology Modeling of Receiver Domain of EIN4. ..... 19
3.4 Comparison of ETR2 and EIN4 Models with ETR1 Receiver Domain Crystal Structure .....24
3.5 Conclusions .....25
4 HOMOLOGY MODELING OF THE KINASE DOMAINS .....26
4.1 Introduction .....26
4.2 Results .....27
4.2.1 Sequence Alignment and Similarity Between Kinase Domains of Ethylene Receptors in Arabidopsis. ..... 28
4.2.2 Predicting Three Dimensional Structure of ETR1 Kinase Domain using Homology Modeling..... 28
4.2.3 Predicting Three Dimensional Structure of ERS1 Kinase Domain Using Homology Modeling..... 31
4.2.4 Predicting Three Dimensional Structure of ETR2 Kinase Domain Using Homology Modeling..... 34
4.2.5 Predicting Three Dimensional Structure of ERS2 Kinase Domain Using Homology Modeling..... 36
4.2.6 Predicting Three Dimensional Structure of EIN4 Kinase Domain Using Homology Modeling..... 38
4.3 Comparison of Models Generated Against 2C2A.A in all the Ethylene Receptors.....40
4.4 Conclusions .....41
5 HOMOLOGY MODELING OF GAF DOMAIN OF ETHYLENE RECEPTORS .....43
5.1 Introduction .....43
5.2 Results .....44
5.2.1 Sequence Alignment and Similarity Between GAF Domains of Ethylene Receptorsin *Arabidopsis*. ..... 44
5.2.2 Predicting Three Dimensional Structure of ETR1 GAF Domain Using Homology Modeling..... 44

5.2.3 Predicting Three Dimensional Structure of ERS1 GAF Domain Using Homology Modeling ..... 46
5.2.4 Predicting Three Dimensional Structure of ETR2 GAF Domain Using Homology Modeling ..... 48
5.2.5 Predicting Three Dimensional Structure of ERS2 GAF Domain Using Homology Modeling ..... 49
5.2.6 Predicting Three Dimensional Structure of EIN4 GAF Domain Using Homology Modeling ..... 51
5.3 Comparison of Models Generated from Common Template 2K2N.A for Subfamily II receptors ..... 52
5.4 Conclusions ..... 53
6 DISCUSSION ..... 54
7 CONCLUSIONS AND FUTURE PROSPECTS ..... 58
REFERENCES ..... 59
APPENDIX ..... 67
VITA ..... 142

# List of Tables

**Table 1**. List of different templates selected by MOE based on Z scores from PDB, for ETR2 receiver domain protein sequence using search function in MOE. ..... 130
**Table 2**. List of different templates selected by MOE based on Z scores from PDB, for EIN4receiver domain protein sequence using search function in MOE. ..... 131
**Table 3**. List of different templates selected by MOE based on Z scores from PDB, for ETR1 kinase domain protein sequence using search function in MOE and PDB BLAST. ..... 132
**Table 4**. List of different templates selected by MOE based on Z scores from PDB, for ERS1 kinase domain protein sequence using search function in MOE and PDB BLAST. ..... 133
**Table 5.** List of different templates selected by MOE based on Z scores from PDB, for ETR2 kinase domain protein sequence using search function in MOE and PDB BLAST. ..... 134
**Table 6.** List of different templates selected by MOE based on Z scores from PDB, for ERS2 kinase domain protein sequence using search function in MOE and PDB BLAST. ..... 135
**Table 7.** List of different templates selected by MOE based on Z scores from PDB, for EIN4 kinase domain protein sequence using search function in MOE and PDB BLAST. ..... 136
**Table 8.** List of different templates selected by MOE based on Z scores from PDB, for ETR1 GAF domain protein sequence using search function in MOE and PDB BLAST. ..... 137
**Table 9.** List of different templates selected by MOE based on Z scores from PDB, for ERS1 GAF domain protein sequence using search function in MOE and PDB BLAST. ..... 138
**Table 10.** List of different templates selected by MOE based on Z scores from PDB, for ETR2 GAF domain protein sequence using search function in MOE and PDB BLAST. ..... 139
**Table 11.** List of different templates selected by MOE based on Z scores from PDB, for ERS2 GAF domain protein sequence using search function in MOE and PDB BLAST. ..... 140
**Table 12.** List of different templates selected by MOE based on Z scores from PDB, for EIN4 GAF domain protein sequence using search function in MOE and PDB BLAST. ..... 141

# List of Figures

**Figure 1.** Structure and classification of five isoforms of Ethylene receptor subfamilies in *Arabidopsis thaliana*. (Lacey and Binder., 2014). ..... 68
**Figure 2.** Role of receiver domain in eliciting ethylene growth responses on etiolated *Arabidopsis* seedlings in the presence of silver ions. Seedlings were grown in darkness for 4 days in air and 100 PPM ethylene on agar plates supplemented with 100µM silver nitrate. Hypocotyl growth of ethylene receptor loss-of-function mutant seedlings was compared to that of wildtype seedlings that were used as controls. *etr1-6;etr2-3;ein4-4* seedlings were transformed with cDNA transgene and the other triple loss-of-function mutant lacked the receiver domain at the end of C terminus. Supplementation and absence of ethylene and silver ions is clearly marked and represented. ..... 69
**Figure 3.** Growth inhibition kinetics of ETR1 receiver domain in the presence of silver ions. The seedlings were grown in air for one hour followed by addition of 1µL $L^{-1}$ ethylene for five hours. Onset of ethylene is indicated by arrow. The seedlings were grown on agar plates containing 100µM silver nitrate. The growth of seedlings was captured every 5 minutes in the time lapse set up and graphs were plotted. The graphs represent time on X-axis and normalized growth rate of hypocotyls on Y-axis. The data represents mean ± SE from at least 4 separate experiments. ..... 70
**Figure 4.** Sequence alignment and identity matrix of receiver domains of Ethylene receptors. . 71
**Figure 5.** Three dimensional structure of ETR2 receiver domain using 1DCF.A as template. A, B. Structurally diverged regions that were conformationally different in the model from respective regions in the template, were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple. ..... 72
**Figure 6.** Structurally diverged regions in target sequence, ETR2 receiver domain in comparison to template, 1DCF.A. ..... 73
**Figure 7.** Three dimensional structure of ETR2 receiver domain using 3RVK.A as template. .. 74
**Figure 8.** Structurally diverged regions in target sequence, ETR2 receiver domain in comparison to template, 3RVK.A. ..... 75
**Figure 9.** Three dimensional structure of ETR2 receiver domain using 1P2F.A. as template. A,B. Structurally diverged regions in the model that were from respective regions in the template, were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple. ..... 76
**Figure 10.** Structurally diverged regions in target sequence, ETR2 receiver domain in comparison to template. 1P2F.A. ..... 77
**Figure 11.** Superimposing the possible models generated for ETR2 receiver domain. ..... 78
**Figure 12.** Three dimensional structure of EIN4 receiver domain using 1DCF.A. as template. . 79
**Figure 13.** Structurally diverged regions in target sequence EIN4 receiver domain in comparison to template, 1DCF.A. ..... 80
**Figure 14.** Three dimensional structure of EIN4 receiver domain using 2R25.B. as template ... 81

**Figure 15.** Structurally diverged regions in target sequence EIN4 receiver domain in comparison to template, 2R25.B. ..... 82
**Figure 16.** Three dimensional structure of EIN4 receiver domain using 3C3M.A as template... 83
**Figure 17.** Structurally diverged regions in target sequence EIN4 receiver domain in comparison to template 3C3M.A..... 84
**Figure 18.** Superimposing the possible models generated for EIN4 receiver domain..... 85
**Figure 19.** Superimposing the possible models generated for ETR2 and EIN4 receiver domain against 1DCF.A..... 86
**Figure 20.** Sequence alignment and identity matrix for kinase domains in Ethylene receptors. Sequence alignment of the kinase domain of the three ethylene receptors in *Arabidopsis* using CLUSTAL W2. The residues with an asterisk (*) denote conserved residues, colons (:) denote conserved substitutions and periods (.) denote semi conserved substitutions (Chenna *et al.,* 2003). B. The identity matrix was created using the same program which shows the percentage of conserved residues among the receptors. ..... 87
**Figure 21.** Three dimensional structure of ETR1 kinase domain using 2C2A.A as template A,B. Structurally diverged regions in the model that were represented in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple..... 88
**Figure 22.** Structurally diverged regions in target sequence, ETR1 kinase domain in comparison to template 2C2A.A. ..... 89
**Figure 23.** Three dimensional structure of ETR1 kinase domain using 3DGE.A as template .... 90
**Figure 24.** Structurally diverged regions in target sequence, ETR1 kinase domain in comparison to template 3DGE.A..... 91
**Figure 25.** Three dimensional structure of ERS1 kinase domain using 2C2A. A as template A,B. Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple. ..... 92
**Figure 26.** Structurally diverged regions in target sequence ERS1 kinase domain in comparison to template 2C2A.A. ..... 93
**Figure 27.** Three dimensional structure of ERS1 kinase domain using 3DGE. A as template A,B. Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple. ..... 94
**Figure 28.** Structurally diverged regions in target sequence ERS1 kinase domain in comparison to template 3DGE.A..... 95
**Figure 29.** Three dimensional structure of ETR2 kinase domain using 2C2A. A as template A,B. Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple. ..... 96
**Figure 30.** Structurally diverged regions in target sequence ETR2 kinase domain in comparison to template 2C2A.A. ..... 97

**Figure 31.** Three dimensional structure of ETR2 kinase domain using 3A0Y. B as template A,B. Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple. ..... 98
**Figure 32.** Structurally diverged regions in target sequence ETR2 kinase domain in comparison to template 3A0Y.B. ..... 99
**Figure 33.** Three dimensional structure of ERS2 kinase domain using 4PL9. A as template A,B. Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple. ..... 100
**Figure 34.** Structurally diverged regions in target sequence ERS2 kinase domain in comparison to template 4PL9.A. ..... 101
**Figure 35.** Three dimensional structure of EIN4 kinase domain using 2C2A. A as template A,B. Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple. ..... 102
**Figure 36.** Structurally diverged regions in target sequence EIN4 kinase domain in comparison to template 2C2A.A. ..... 103
**Figure 37.** Three dimensional structure of EIN4 kinase domain using 4Q20.A as template A,B. Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple. ..... 104
**Figure 38.** Structurally diverged regions in target sequence EIN4 kinase domain in comparison to template 4Q20.A. ..... 105
**Figure 39.** Comparison of models generated against 2C2A. A for all receptors except ERS2. The models with least RMSD value generated against 2C2A. A were superimposed in MOE. Panel A represents structurally conserved alpha helices in DHp and CA subunits of kinase domain. Panel B represents structurally diverge regions mainly constituting beta strands of CA subunit. In both the panels, ETR1 was represented in orange, ERS1 in purple, ETR2 in cyan, EIN4 in light brown and crystal structure of the template, 2C2A. A in dark brown colors. ..... 106
**Figure 40.** Structurally diverged/similar regions in target sequences in comparison to template. The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged/similar regions were selected and represented in sequence Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodels for all the receptors except ERS2 and template as 2C2N. A. The top most panel represents gap in DHp regions corresponding to loop formation in the secondary structure. The middle panel represents the most conserved region among all the receptors. The lower panel represents the most diverge regions on the receptors in comparison to each other as well as with the template. ..... 107
**Figure 41.** Sequence alignment and identity matrix of GAF domains of Ethylene receptors. .. 108

**Figure 42.** Three dimensional structure of ETR1 GAF domain using 1UI6.A as template..... 109
**Figure 43.** Structurally diverged regions in target sequence ETR1 GAF domain in comparison to template 1UI6.A. ..... 110
**Figure 44.** Three dimensional structure of ETR1 GAF domain using 1B6A.A as template ..... 111
**Figure 45.** Structurally similar/conserved regions in target sequence ETR1 GAF domain in comparison to template 1B6A.A. ..... 112
**Figure 46.** Three dimensional structure of ERS1 GAF domain using 3TTG. A as template A,B. Structurally conserved regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple. ..... 113
**Figure 47.** Structurally similar/conserved regions in target sequence ERS1 GAF domain in comparison to template 3TTG.A. ..... 114
**Figure 48.** Three dimensional structure of ERS1 GAF domain using 3FY4.A as template A,B. Structurally conserved regions in the model that were conformationally similar from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple. ..... 115
**Figure 49.** Structurally similar/conserved regions in target sequence ERS1 GAF domain in comparison to template, 3FY4.A. ..... 116
**Figure 50.** Three dimensional structure of ETR2 GAF domain using 2K2N.A as template A,B. Structurally conserved regions in the model that were conformationally similar from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple. ..... 117
**Figure 51.** Structurally similar/conserved regions in target sequence ETR2 GAF domain in comparison to template 2K2N.A. ..... 118
**Figure 52.** Three dimensional structure of ETR2 GAF domain using 4HL7.A as template ..... 119
**Figure 53.** Three dimensional structure of ERS2 GAF domain using 2K2N. A as template A,B. Structurally conserved regions in the model that were conformationally similar to respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple. ..... 120
**Figure 54.** Structurally similar/conserved regions in target sequence ERS2 GAF domain in comparison to template 2K2N.A. ..... 121
**Figure 55.** Three dimensional structure of ERS2 GAF domain using 2K7W.A as template A,B. Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple. ..... 122
**Figure 56.** Structurally diverge regions in target sequence ERS2 GAF domain in comparison to template 2K7W.A. ..... 123
**Figure 57.** Three dimensional structure of EIN4 GAF domain using 2K2N.A as template. A,B. Structurally diverged regions in the model that were conformationally different from respective

regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple. ..... 124
**Figure 58.** Structurally diverge regions in target sequence EIN4 GAF domain in comparison to template 2K2N.A. ..... 125
**Figure 59.** Three dimensional structure of EIN4 GAF domain using 2PRR.A as template A,B. Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple. ..... 126
**Figure 60.** Structurally diverge regions in target sequence EIN4 GAF domain in comparison to template 2PRR.A. ..... 127
**Figure 61.** Superimposing three dimensional structures of subfamily II GAF domains using 2K2N.A as template ..... 128
**Figure 62.** Structurally similar regions in target sequences of GAF domains of subfamily II receptors in comparison to template 2K2N.A. ..... 129

# 1 INTRODUCTION

## 1.1 Ethylene as a Phytohormone

Throughout the length of life, plants maintain highly plastic growth by adapting to changes in the environment which involve many physiological and anatomical changes. They are exposed to a variety of biotic and abiotic stresses and exhibit complex responses to these stress stimuli. Such responses are mediated by small endogeneous molecules known as phytohormones. These play vital roles in mediating growth and development in plants. Research by early plant biologists led to the discovery of five phytohormones in plants, considered as classic phytohormones. They are well characterized and include auxin, cytokinin, abscisic acid, gibberellic acid and ethylene. Recent studies led to discoveries that added more chemicals to this list, which include salicylic acid, polyamines, brassinosteroids, jasmonic acid, nitric oxide, strigolactones and peptide hormones (Santner., 2009).

Ethylene is a simple molecule – $C_2H_4$, an unsaturated hydrocarbon with two carbons connected by a double bond, and four hydrogens. It was once commonly used as an anesthetic along with oxygenin for surgical medicine (Fairlie., 1929). Horticulturally, it is mainly used in fruit ripening and for getting produce to consumers without spoiling or over-ripening. Chemically, ethylene is a simple molecule but functionally, it is very complex. Ethylene is diffusible and has different roles in plant growth and development such as fruit ripening, seed germination, seedling growth, flowering, leaf and organ abscission, senescence, gravitropism, responses to biotic and abiotic stresses, nutations and triple responses (exaggerated apical hook, elongated roots, radial swelling and inhibition on hypocotyl) (Abeles., 1992; Binder., 2006, and Berg., 1992)

## 1.2 Ethylene Receptors in *Arabidopsis thaliana*

In the model organism *Arabidopsis thaliana*, a family of five disulfide-linked homodimer receptors mediate ethylene responses. These receptor isoforms bind to ethylene with high affinity and are divided into two subfamilies based on their sequence comparisons: Subfamily I has ETR1 (Ethylene Response 1), ERS1 (Ethylene Response Sensor1) and subfamily II has ETR2, ERS2 and EIN4 (Ethylene Insensitive 4). Although they are functionally diverged, these receptor isoforms share several features in common (Fig 1) (All the figures and tables are listed at the end of the document). The N-terminal of all these receptor isoforms have three alpha helices that are embedded in the membrane of endoplasmic reticulum and form the ethylene-binding domain. This is followed by a GAF domain (cGMP-specific phosphodiesterases, adenylyl cyclases, and FhlA) that is suggested to perform receptor-receptor interactions.

The kinase domain is located after the GAF domain. The subfamily I receptors contain histidine kinase activity and subfamily II receptors contain a degenerate histidine kinase activity and show serine/threonine kinase activity. ERS1 has both histidine kinase and serine/threonine kinase activity. A subset of ethylene receptors, ETR1, ETR2 and EIN4, contain a receiver domain that has a conserved aspartate residue, towards the C-terminal region. The GAF domain, kinase domain and the receiver domain are responsible for the signal output and are homologous to the bacterial two-component system which function with a phospho-relay mediated signal transduction mechanism.

## 1.3 Ethylene Signal Transduction Pathway

The first step in ethylene signal transduction pathway is binding of ethylene to the receptors. It is thought to negatively regulate these receptors. Upon ethylene binding, the kinase transmitter domain of members of the receptor family interacts with the regulatory domain of the Raf-like

kinase CTR1 and inactivates it. CTR1 negatively regulates EIN2. When ethylene is bound, this negative regulation is removed. The cytoplasmic C-terminal domain of EIN2 is proteolytically processed such that it positively signals downstream to the EIN3 family of transcription factors located in the nucleus. A target of the EIN3 transcriptions factors is the promoter of the ERF1 gene, a member of a second family of transcription factors, thus initiating transcriptional response to ethylene.

## 1.4 Role of Metal Ions in Ethylene Receptor Function and Signaling

Metal ions play an important role in ethylene receptor function. In 1999, Rodriguez *et* al., suggested that metal ions are important in binding of ethylene to the receptors. Copper, is a metal cofactor required for ethylene binding to the receptors. Also, it has been reported that copper is also important in the biogenesis of these receptors. Silver ions block ethylene perception in plants, but support ethylene binding to ETR1. Binding assays were performed on truncated ethylene binding domains of the receptors that were constitutively expressed in *Pichia pastoris.* It was reported that copper and silver support ethylene binding to subfamily I receptors but only copper supports binding to subfamily II receptors. This suggested that there is differential binding of ethylene in the presence of silver between the two sub families. (Rodriguez., 1999, McDaniels and Binder, 2012).

## 1.5 Effect of Silver Nitrate on Ethylene Growth Responses of Dark-Grown *Arabidopsis* Seedlings

McDaniels and Binder also looked at overlapping functions of ethylene receptors in the presence of Silver ions. In the wildtype backgrounds of the various receptor mutants (Col, WS), $AgNO_3$ blocked growth inhibition caused by addition of ethylene. All single Loss-of-function (LOF) receptor mutants exhibited insensitivity to ethylene in the presence of $AgNO_3$ with the exception

of *etr1-7* and *etr1-9*. The triple LOF mutant *etr1-6;etr2-3;ein4-4* exhibited a constitutive ethylene response in air and growth was inhibited further by application of ethylene in the presence of silver.

Generally, there are two phases of growth inhibition. The first phase starts approximately in 10 minutes after ethylene is introduced and reaches a plateau after around 10 minutes and lasts for about 30 minutes. The second phase of growth inhibition follows the plateau and persists as long as the ethylene is present. The effect of silver nitrate on ethylene growth response kinetics was studies by McDaniels and Binder in 2012. In the absence of $AgNO_3$, wild type seedlings exhibited the two defined phases of growth inhibition; however, with the exception of a short, transient response, the presence of 100 µM[milli molar] $AgNO_3$ abolished both phases. In the triple loss-of-function mutant, a two-phase growth inhibition response occurred, but there was no reversal of growth inhibition observed in the presence of 100 µM $AgNO_3$.

To identify the individual roles of each receptor isoform in mediating silver's effect, each receptor isoform was transformed into the *etr1-6;etr2-3;ein4-4* under the control of the ETR1 promoter. ETR1 was the only transgene that could fully rescue the effects of silver where as ERS1, ERS2 and EIN4 elicited transient responses. The only transgene unable to complement silver's inhibitory effect was ETR2. This suggested that ETR2 has no apparent role in response to Silver. This did not correlate well with their binding assays data, suggesting that there is a second metal ion binding site elsewhere on the receptor apart from the ethylene binding domain.

## 1.6 Sub Functionalization of Ethylene Receptors

Although the ethylene receptors are very similar in their structures, they have diverse functions with both overlapping and non-overlapping roles. Previous data by Wilson *et al.*, in 2014 suggested some functions that led to the sub functionalization of ethylene receptors. They

proved that ETR1 and ETR2 have contrasting roles in seed germination under salt stress. ETR1 and EIN4 inhibit seed germination under salt stress. ETR2 stimulates seed germination under salt stress. Full length ETR1 is required for ethylene-stimulated nutations in dark grown Arabidopsis seedlings. ETR1, ETR2 and EIN4 play an important role in growth recovery after removal of ethylene (Wilson *et al.*, 2014 a, b). Hence, we wanted to look at the structures of these domains for a better understanding of their physiological roles. Unfortunately, there is not much information available in the protein data bank regarding the crystal structures of these domains , except for the receiver domain of ETR1 and a part of the kinase domain of ETR1. Hence, we resorted to knowledge-based three dimensional structure prediction of these proteins, using homology modeling.

## 1.7 Molecular Modeling Using Computational Approaches

Functional characterization of a protein sequence is of paramount importance to properly understand the activity or functioning of the protein. Research so far suggests only less than one percent of the sequences in the fast growing current sequence databases have been experimentally verified (UniProt Consoritum). Although structural biologists world-wide have been substantially contributing to this, this situation is very unlikely to change. The limited progress towards the challenge of structure prediction has been daunting to experimentalists. The only viable solution available to this complex question is the invention of automated computational approaches that effectively compute protein function on a variety of characterstics such as sequence similarity, evolutionary relationship, presence of common motifs etc. (Honig *et al.*, 2015).

Understanding the molecular details of a protein is of paramount importance in the field of scientific research and with the intervention of computational approaches, the progress in the

past few years has been incremental. Faster computers, availability of information regarding the sequence and evolutionary relationships of proteins, bioinformatics algorithms have been able to predict structural models of a plethora of sequences with reasonable accuracy. These approaches not only help visualize the three dimensional structure of proteins, but also are used to give better insights into the low resolution images that were generated by techniques such as electron microscopy, protein-protein interactions, ligand binding etc (Russell *et al*., 2005)

## 1.8 Homology Modeling of Proteins

Homology modeling of proteins is an automated comparative modeling approach that is mainly dependent on making use of sequence similarities and identities between proteins to generate three dimensional modeling of proteins whose structures are unknown, using their homologues as templates. (Greer 1981, Blundell 1987). Homology modeling approaches have improved over the past decade, but the protocol employed for structure predictions, has by and large remained unchanged with a few exceptions. The step-wise approach to homology modeling consists of identification of right template(s) to the query sequence, structure preparation of the selected template(s), sequence alignment of the template(s) selected, modeling of the conserved regions and regions that are structurally divergent (SDR), and refinement and analysis of the generated models.

## 1.9 Goal of the Current Study

The main goal of the current study is to predict the three dimensional structures of the different cytosolic domains of ethylene receptors to obtain further insights into the sub functionalization of the receptors. We could predict the approximate structures of the same and compared each of the predicted structures of each domain in all the receptor isoforms for a better understanding of how conformational changes in the structure result in different physiological functions.

# 2 MATERIALS AND METHODS

## 2.1 Seedling Preparation

*Arabidopsis thaliana* seeds were subjected to surface sterilization by treating them in 70% alcohol for 30 to 60 seconds. The seeds were then imbibed in about 100µl of distilled water in a microfuge tube wrapped in aluminum foil, for dark cold treatment at 4ºC. After stratification for 3 to 4 days, they were light treated for 4 to 8 h under continuous fluorescence lights. Following light treatment, the seeds were carefully placed on agar plates containing one-half strength Murashige and Skoog basal salt mixture (Murashige and Skoog., 1962) , pH 5.7, 0.8% (w/v) agar, consisting of inorganic salts: $NH_4NO_3$, 400; KCI, 65; $KNO_3$. 80; $KH_2PO_4$, 12.5; $Ca(NO_3)_2 \cdot 4H_2O$,144; $MgS0_4$. $7H_2O$. 72; NaFe-EDTA, 25; $H_3BO_3$, 1.6; $MnS0_4 \cdot 4H_2O$, 6.5; $ZnSO_4 \cdot 7H_2O$, 2.7: and KI. 0.75 and organic substances including inositol (100 mg $mL^{-1}$), nicotinic acid (1 mg $mL^{-1}$), pyridoxin HCl (1 mg $mL^{-1}$), and thiamine HCl (10 mg $mL^{-1}$) with no added sugar. For experiments with silver treatment, 100µM of silver nitrate ($AgNO_3$) was added to the agar before pouring the plates and after the media is cooled at 65ºC. These plates were always wrapped in Aluminum foil to avoid exposure to light.

The mutant seed lines *etr1-6, etr1-7, etr2-3, ers2-33* and *ein4-4* were obtained from the Meyerowitz lab, *ers 1-3* and *etr1-9* from the Schaller lab. The *etr1-6, etr1-7, etr2-3, ers2-3* and *ein4-4* are in Columbia (Col) background and *etr1-9, ers1-3* and *ers1-2* are in Wassileweskija (Ws) background.

## 2.2 Growth-Rate Measurements of Hypocotyls

### 2.2.1 End Point Analysis

The agar plates containing the seeds were carefully wrapped with aluminum foil, to maintain the seedling growth in darkness. The plates were vertically aligned in gas-tight chambers with a

continuous flow of 100 ppm ethylene or hydrocarbon-free air at a flow rate of 100mL $min^{-1}$. The seedlings were allowed to grow for 4 days in darkness at room temperature, 22ºC to 25ºC. The plates were then scanned on a flat-bed scanner and the length of the hypocotyls was measured using ImageJ (ver 1.48; http://imagej.nih.gov/ij/)

### 2.2.2 High Resolution Time Lapse Imaging

The seedlings were grown on vertically oriented plates in darkness to a hypocotyls length of 2 to 4 mm at room temperature, 22ºC to 25ºC for measuring the growth-rate. The plates with the seedlings were fitted in a chamber that allows continuous gas flow and mounted on a micromanipulator to keep in position through the entire length of the experiment. The parameters were set to grow the seedlings in one hour of ethylene-free air followed by five hours of 1 ppm ethylene. A continuous gas flow of 100mL $min^{-1}$ was maintained throughout the experiment (Binder, 2004b). All the experiments were performed in the dark to avoid exposure of seedlings to light.

Electronic images of the hypocotyls growth patterns were captured every 5 minutes using a computer driven, charge - coupled device (CCD) camera equipped with a close-focus zoom lens (Spalding et al., 1998). Image resolution was maintained between 120 and 150 pixels per millimeter. The height of the seedlings was measured in pixels, in each frame and the growth rates were calculated by Lab VIEW Environment (National Instruments) using custom software (Spalding *et al.,* 1998) and was normalized to the growth rate in air for the first hour, prior to addition of ethylene.

## 2.3 Homology Modeling of Ethylene Receptors

The three dimensional structural models of different cytosolic domains of ethylene were generated with a sequence-to-structure prediction approach using Molecular Operating

Environment, version 2012.10 (MOE - http://www.chemcomp.com/MOE-Protein_and_Antibody_Modeling.htm). Generation of each model involved the following steps.

### 2.3.1 Template Identification

Sequence based template selection was done using two approaches. In one, the query sequence was fed into the search tool in MOE window and was commanded to look for templates with highest Z values. The higher the Z value, the closer are the sequences structurally.

The second approach was to use PDB BLAST (Protein Data Bank Basic Local Alignment Search Tool) to look for templates in PDB that have higher sequence similarity and lowest e-values. The templates were selected and their sequences were downloaded from PDB in the form of a ".pdb" file and were opened in MOE for further modeling.

### 2.3.2 Structure Preparation

The query sequence and template were loaded into MOE and the structures were prepared and corrected to fix the errors automatically. The hydrogen bond network was optimized by using Protonate 3D function which calculates optimal protonation states for the structure. (Labute, 2008)

### 2.3.3 Amino Acid Sequence Alignment and Correction

Both the sequences were aligned initially to check the BLOSUM scores between them and to minimize gap penalty, a range of BLOSUM matrices were tried from BLOSUM30 to BLOSUM62 and the sequence similarity and identity scores were calculated. The BLOSUM matrix with least gap penalty and high sequence similarity was selected and set for model generation.

### 2.3.4 Backbone Generation

The model backbone of the model was generated using CHARMM27 (Chemistry at HARvard Moelcular Mechanics) force field that allows a wide range of simulations while generating the structure of protein using the template sequence. An ensemble of 10 possible structures was created and an optimized and energy minimized 11th structure was finally generated averaging the previously generated structures.

# 3 RECEIVER DOMAIN

## 3.1 Introduction

A subset of ethylene receptors, ETR1, ETR2 and EIN4 in *Arabidopsis* contain receiver domain at their C-termini. Receiver domains have been identified in both the receptor subfamilies in dicots but only in subfamily II members in monocots (Binder *et al* ., 2012). Receiver domains are homologous to response regulators, which are important domains in the two-component signal transduction system in bacteria. These primarily act as phosphorylation switches and signal output domains. The receiver domain is proven to have an important role in ethylene -stimulated nutations (Binder *et al.,* 2006) and in the inhibitory role of ETR1 on seed germination during salt stress (Kim *et al.,* 2011; Wilson *et al.,* 2014b). ETR1 is a hybrid two-component system with the transmitter and receiver in the same molecule. Receiver domains mainly function through phosphotransfer. However, it was shown that phosphotransfer through the ETR1 receiver domain is not always necessary for responses to ethylene. Such kinase-independent roles include nutational bending of dark-grown *Arabidopsis* seedlings- stimulated by ethylene (Binder *et al.,* 2006).

Silver ions act as non-competitive inhibitors to ethylene and inhibit its perception in plants (Beyer., 1976) Research by McDaniels and Binder in 2012 studied different effects of silver nitrate on ethylene growth responses of dark-grown *Arabidopsis* seedlings. They suggested that ETR1 is important for responses to silver since *etr1-6* loss-of-function mutants had little or no response to silver and triple *etr1-6;etr2-3;ein4-4* mutants had no response to silver ions. This triple loss-of-function mutant has functional ERS1 and ERS2 that do not have receiver domains. They also suggested that phosphotransfer through ETR1 or a functional RTE1 is a mandatory requirement for the effects of silver nitrate in the presence of ethylene. These observations point

to the fact that receiver domain of ETR1 is not an absolute requirement to elicit responses to ethylene in the presence of silver ions.

## 3.2 Role of ETR1 Receiver Domain in Responses to Silver

### 3.2.1 Role of Receiver Domain in Ethylene Blocking Effects of Silver Nitrate

To more completely evaluate the role of the ETR1 receiver domain in mediating responses to silver ions in the presence and absence of ethylene, wild type and mutant seedlings were grown on 1/2 MSNS agar plates, supplemented with 100µM silver nitrate, for 4 days in the dark in air and 100 PPM ethylene continuously flowing. Consistent with the previous studies by McDaniels and Binder in 2012, in wild type, the growth of the etiolated seedlings was unaffected by ethylene in the presence of silver ions. When the triple mutants were transformed with a full length ETR1 transgene (cETR1) or a truncated transgene lacking receiver domain (*cetr1* - ΔR), they exhibited insensitivity to ethylene in the presence of silver ions. Contrast to these observations, the triple loss-of-function mutant exhibited an ethylene response in the presence of silver (Fig. 2). Together, this suggests that the ETR1 receiver domain is not required for responses to silver ions.

To confirm these results, time lapse imaging was done on dark-grown *Arabidopsis* seedlings, that were grown on agar plates supplemented with 100µM $AgNO_3$, to study the growth kinetics in the presence of 1PPM ethylene gas. There are two phases of growth inhibition. The first phase starts approximately 10 minutes after ethylene is introduced and reaches a plateau after around 10 minutes and lasts for about 30 minutes. The second phase of growth inhibition follows the plateau and persists as long as the ethylene is present. Consistent with a previous study (McDaniels and Binder, 2012), in the presence of silver, the wild-type seedlings, showed no response to ethylene and the triple loss-of-function mutant seedlings showed the two phases of

growth inhibition. When the triple mutants were transformed with a full length ETR1 transgene (cETR1) or a truncated transgene lacking receiver domain (*cetr1* - ΔR), the seedlings showed a small and transient growth inhibition response on application of ethylene. Both the transformants had receiver domain but could not elicit a long term response to ethylene in the presence of silver ions (Fig 3). This suggests that ETR1 receiver domain has little or no role in mediating ETR1 responses to silver ions.

## 3.3 Homology Modeling of Cytosolic Domain of Ethylene Receptors

The hypothesis tested above suggested that the receiver domain does not play an important role in the effects of silver ions in response to ethylene. However, it is known to be important for other responses (Wilson *et al.,* 2014). This led to an interest in studying the structural aspects that lead to the sub-functionalization of ethylene receptors. This can be achieved by visualizing the structure of the cytosolic domains of the ethylene receptors in *Arabidopsis*. Unfortunately, the information regarding the crystal structure of different domains of ethylene receptors is only limited to the receiver domain of ETR1 (PDB ID: 1DCF, Muller-Dieckmann *et al.,* 1999) and catalytic ATP-binding domain of ETR1 (Mayerhoff *et al.,* 2014). To understand the structural aspects in detail, we resorted to predicting the three dimensional protein structure using knowledge-based prediction - homology modeling.

The large cytoplasmic domain of the ethylene receptors has significant sequence homology to the two-component signal transduction system in bacteria (Chang *et al.,* 1993). The receiver domain at the C terminus of these receptors is only present in a subset of ethylene receptors, in ETR1, ETR2 and EIN4. The receiver domain of ETR1 in *Arabidopsis* was crystallized and showed to have high conservation with bacterial receiver domains despite lower sequence similarity (Muller - Dieckmann *et al*., 1999). It was shown that the orientation of γ-loop in

ETR1, which is involved in molecular recognition is in a different confirmation from the other proteins for different functions of the ETR1 receiver domain. The structure of a receiver domain typically consists of five α-helices and five β sheets alternating with each other. It has six highly conserved residues - three aspartic residues forming the acidic pocket, and along with a lysine to coordinate a metal ion cofactor in the acidic pocket. One of the aspartates in the acidic pocket is the site of phosphorylation. The other two residues are a serine/threonine and phenylalanine/tyrosine that are involved in conformational changes and signal output (Bourret, 2010). In ETR1, the length of receiver domain is about 124-128 residues and is a dimer in solution and crystal (Muller-Dieckmann *et al.,* 1999).

### 3.3.1 Sequence Alignment and Similarity Between Receiver Domains of ETR1, ETR2 and EIN4.

Protein sequences that are evolutionarily related have similar structures measured by their sequence similarity. The branch length between two protein nodes in a phylogenetic tree is measured by their sequence identity, which defines the position of the two proteins representing their evolutionary relationships. We used CLUSTALW2 (an improved version of CLUSTAL W), a multiple sequence alignment tool for divergent proteins (Thompson *et al.,* 1994), to compare the sequence identity and similarity between the receiver domains of ETR1, ETR2 and EIN4 (Fig 4 A,B).

The residues indicated with an asterisk (*) denote identical residues in all sequences that are conserved, colons (**:**) denote highly conserved substitutions and periods (**.**) denote semi conserved substitutions indicating weakly conserved residues (Chenna *et al.,*2003). The percentage identity matrix created from CLUSTAL W suggested that receptors in the subfamily II, ETR2 and EIN4 are more closely related to each other than to ETR1 (Fig 4B).

### 3.3.2 Predicting Three Dimensional Structure of ETR2 Receiver Domain

Proteins with similar or related sequences adopt similar structures. Comparative modeling using knowledge-based prediction generates protein structures to an approximation by conforming to homologous atoms and special constraints. We hereby refer the sequence of the protein whose structure is unknown as 'query' or 'target' and the sequence of the protein with known structure as 'template'. Previous studies suggest that the sequence identity between the template and the query should be around 30% to obtain a fairly accurate protein structure prediction. We used Molecular Operating Environment (MOE 2012) to generate the three dimensional structures of the target proteins.

ETR2 belongs to subfamily II of the ethylene receptors and its receiver domain spans residues from 647 - 773. A list of target templates were selected using PDB from MOE, based on Z value. The query was modeled against ten templates that were chosen by MOE based on highest Z scores (Table 1). It was clearly evident from the list of the templates that all of them were response regulators in bacteria and other lower organisms and were homologous to the ethylene receptor receiver domain in plants. The sequence identity and similarity of the templates suggest that 1DCF.A has the highest sequence identity of 34.6 and similarity of 55.9 with the query sequence. The RMSD values were tabulated for comparative analysis. It was interesting to note that the templates with higher sequence identity/similarity compared to others did not always yield models with commensurate lower RMSD values. The models were also aligned to the crystal structure in PyMol and the values were tabulated. PyMol helps in aligning the structures based on the position of their α-carbons and hence was chosen to have a better alignment of the backbone carbon atoms.

Graphs were plotted with PDB ID on X-axis and RMSD values on Y-axis for both MOE and PyMol superimposed structures. The average RMSD value of the templates superimposed in

MOE was around 0.93Å and that of PyMol was approximately 2.2Å [Angstrom]. The reason for such high RMSD value was unknown as alignment of the models against the templates 1W25.A and 2AYX.A resulted in unusually high RMSD values in spite of aligning the backbone carbons on the top of each other. Overall comparison of sequence identity/similarity and RMSD values suggest that models generated against the templates 1DCF.A, 1P2F.A and 3RVK.A would be approximately closer to the three dimensional structure of ETR2 receiver domain.

### 1.1.1 A 3D Model for Structure of ETR2 Receiver Domain Using 1DCF. A as Template

1DCF.A is the PDB ID for crystal structure of the receiver domain of the ETR1 in *Arabidopsis thaliana*. The model with lowest RMSD among all the models generated against 1DCF.A was the 4th model with an value of 0.89Å when superimposed in MOE and 0.121Å when superimposed in PyMol. The predicted secondary structure closely resembled the crustal structure 1DCF.A, receiver domain of ETR1, consisting of a beta strand at the N terminus, followed by five alpha helices and four more beta strands, all alternating with each other. The components of secondary structure were connected by loops and turns

It was observed that the target sequence could not build a complete secondary structure and shows a short alpha helix (Fig 5A) due to a major gap in the template from residue 45 - 51, annotated as 'RVV - - - - - - - SHEH' aligned against the region from residue 37 - 49, represented as 'TAIAPGSSSPSTS'. The region that is not aligned well is modeled as a loop in this region. (Fig 5B, 6). Similarly, the target could not fill the gap in the region from 72 - 78, 'RSR - - - - - SWPL' aligned against the template with sequence 'HEKFTQRHQRPL' from residues 73-85 and hence generated a loop (Fig 5B, 6).

#### 3.3.2.1 A 3D Model for Structure of ETR2 Receiver Domain Using 3RVK.A as Template

3RVK.A is the PDB ID for structure of the CheY-Mn2+ Complex with substitutions at 59 and 89: N59D E89Q in *Escherichia coli*. The model with lowest RMSD among all the models generated against 3RVK.A was the 6$^{th}$ model with an value of 0.58Å when superimposed in MOE and 1$^{st}$ model with 0.173Å and that of 6$^{th}$ model was 0.234Å when superimposed in PyMol. The secondary structures closely resembles 1DCF.A but the fifth alpha helix is placed towards the end of the protein while there was a tail of sequence at the end of 1DCF.A that is more likely to be a beta strands (Mueller-Dieckmann*et al.*, 1999). (Fig 7).

There is a minor gap in the between the first alpha helix and the second beta strand in the orientation of the loop due to a gap generated in the target sequence from residues 24 - 26 denoted by 'CD - V' aligned against residues 24 -27 'FNNV' in the template (Fig 7B,8). There is a major difference in the loop region between second alpha helix and third beta strand as the target majorly diverged from the template's loop alignment from residues 41 - 52, 'PGSSSPSTSFQV' as there is a gap at that region around residues 42 - 47, 'AGG - - - - - - YGF' (Fig 7A,8).

At the end of the third alpha helix and the loop connecting it with the fourth beta strand, there is a slight divergence in the target sequence as the alignment generated a gap around residues 73 - 76, 'SR - - - SW' against the region 'ADGAMSA' in the template sequence from residues 68 - 74 (Fig 7B,8).

#### 3.3.2.2 A 3D Model for Structure of ETR2 Receiver Domain Using 1P2F.A as Template

1P2F.A is the PDB ID for crystal structure of response regulator DrrB, a *Thermotogamaritima* OmpR/PhoB Homolog. The model with lowest RMSD among all the models generated against 1P2F.A was the 5$^{th}$ model with an value of 0.37Å when superimposed in MOE and 9$^{th}$ model

with 0.181Å and that of 5$^{th}$ model was 0.215Å when superimposed in PyMol. The secondary structure is similar to 1DCF.A but slightly different with the second alpha helix is short with only one turn compared to other alpha helices (Fig 9A).

A slight divergence at the end of the first alpha helix due to a gap generated in the target sequence around the residues 22-24 'L - - GC' against the residues between 22 - 26, 'LQQLG'. Here, the beta strand with a Glycine, G, begins earlier in the target sequence than the template sequence (Fig 9B,10). The sequence right after a short second alpha helix in the template from 39 - 45 denoted as 'ND - - - - - - - - EEAFH' is aligned against a long looped region in the target sequence from 37 - 51, 'TAIAPGSSSPSTSFQ' (Fig 9 A,10). A slight break between the residues R72 and S73 represented as 'R-S' in the target sequence created a loop in the place of a turn in the third alpha helix 'KET' between 67 and 68 in the template (Fig 9B,10).

#### 3.3.2.3 Comparison of the Three Predicted Structures of ETR2

The models generated using the templates with lowest RMSD values when superimposed against the respective templates, 1DCF.A**,** 3RVK.A and 1P2F. A were superimposed against each other in PyMol to look at the regions that are diverged among the three models and the results were tabulated (Data not shown) (Fig 11).

The structures significantly have the loop regions diverged from others all throughout. The basic blueprint of the tertiary structure remained similar for most of the protein, but with some exceptions in the lengths an orientations of the backbone carbons. The second beta strand was shorter in 1DCF.A and 3RVK.A when compared to 1P2F.A. The second alpha helix extremely diverged in all the proteins, with a single turn and double turn in 1P2F.A and 3RVK.A respectively compared to 1DCF.A which has three turns and the loop orientation is at different angles in all the three models (Fig 11A).

It was interesting to note that the γ-loop which is involved in various key physiological functions of the receptor was oriented similarly in 3RVK.A and 1P2F.A but was at an angle of about 150° in case of 1DCF.A. Towards the end of the protein models, the structures were different in the length and orientation of the end of the fifth alpha helix and the stretch of the tail sequence. (Fig 11B, 12)

### 3.3.3 Homology Modeling of Receiver Domain of EIN4.

EIN4 belongs to subfamily II of the ethylene receptors and its receiver domain spans the residues from 643 - 766. The templates were selected (Table 2) based on Z s cores and models were generated against each template, as described above.

Graphs were plotted with PDB ID on X-axis and RMSD values on Y-axis for both MOE and PyMol superimposed structures. The average RMSD value of the templates superimposed in MOE was around 0.71Å and that of PyMol was approximately 1.38Å . The reason for high RMSD values for PyMol alignment of 2AYX.A and 3R0J.A was unknown. Overall comparison of sequence identity/similarity and RMSD values suggest that models generated against the templates 1DCF.A, 2R25.B and 3C3M.A would be approximately closer to the three dimensional structure of EIN4 receiver domain.

#### 3.3.3.1 A 3D Model for EIN4 Receiver Domain Using 1DCF.A as Template

The model with lowest RMSD among all the models generated against 1DCF.A was the 4$^{th}$ model with an RMSD value of 0.33Å when superimposed in MOE and 5$^{th}$ model 0.13Å when superimposed in PyMol. The predicted secondary structure closely resembled the crystal structure 1DCF.A, receiver domain of ETR1, consisting of a beta strand at the N terminus, followed by five alpha helices and four more beta strands, all alternating with each other. The components of secondary structure were connected by loops and turns. The sequence identity

and similarity of the query sequence with the template were 37.9 and 64.5 respectively and were aligned using BLOSUM45.

Figure 12A shows significantly diverged regions and possible orientations of the diverged regions of different models generated against 1DCF and superimposed the same with the template crystal structure. Major divergences are in the loop connecting the second alpha helix to the third beta strands which spans the sequence 'ALSNVEMSYR' in the query from residues 38 to 47 in the 4$^{th}$ model with least RMSD value. This represents a gap in the template sequence from residue 46 - 52 with the sequence 'VVSH - - - EHK' but considers a possibility of a small alpha helix in the predicted secondary structure of the query sequence by MOE (Fig 12C, 13).

The sequence between the third alpha helix and fourth beta strands with residues from 66- 75 in the query sequence (KI - RKF - CGHHW) significantly diverged against the sequence 'RIHEKFTQRHQR' in the template from residues 71-83(Fig 12D, 13). The software predicted that the loop started earlier, with a phenylalanine but the same phenylalanine was represented in the alpha helix in the template crystal structure.

Towards the end of C-terminal, there was a difference in the orientation of the generated model. The sequence spanned between the residues 116 - 125 (RRAL - - QTASE) produced a turn against the template sequence ' SDLLEPRVLYE' between the residues 124-135 (Fig 12D, 13).

#### 3.3.3.2 A 3D Model for EIN4 Receiver Domain Using 2R25.B as Template

2R25.B is the PDB ID for crystal structure of complex of YPD1 and SLN1-R1 with bound Mg2+ and BeF3- in *Saccharomyces cerevisiae*. The model with lowest RMSD among all the models generated against 2R25.B was the 1$^{st}$ model with an RMSD value of 0.54Å when superimposed in MOE and 1$^{st}$ model with 0.174Å when superimposed in PyMol. The predicted secondary structure closely resembled the crystal structure 2R25.B, receiver domain of ETR1, with slight

variations in their lengths, consisting of a beta strand at the N terminus, followed by five alpha helices and four more beta strands, all alternating with each other. The components of secondary structure were connected by loops and turns. The sequence identity and similarity of the query sequence with the template were 33.1 and 56.5 respectively and were aligned using BLOSUM35. Figure 14A shows significantly diverged regions and possible orientations of the diverged regions of different models generated against 2R25.B and superimposed the same with the template crystal structure. Structural divergences were observed in the second beta strand and loop connecting it to the second alpha helix to the third beta strand which spanned the sequence 'TAVS - - - - SGF' in the query from residues 27 to 33 in the 5$^{th}$ model with the least RMSD value. This represents the amino acids from residue 1115 - 1125 in the template with the sequence 'IELACDGQEAF'. The model generated by MOE has a shorter alpha helix and modeled a loop to fill in the residues in the gap compared to a slightly longer alpha helix in the template crystal structure (Fig 14C, 15).

A short stretch of amino acids between 40 -43 in the query could not model an alpha helix for the first two residues in the sequence 'SNVE' against 'TSKG' of the template sequence between 1132-1135 which is the second alpha helix extended through the first three residues and a turn at the position of Glycine (1135). The end of the loop between the third alpha helix and fourth beta strand with residues from 74- 79 in the query sequence (HWPLII) diverged against the sequence 'TSP - IV' in the template from residues 1166-1170 (Fig 14D, 15). The gap after the Proline in the query sequence produced a wider turn compared to that of the crystal structure.

The model significantly diverged from the template between the fifth beta strand and the fifth alpha helix in the sequence spanned between the residues 102-111 (IQKPVL - - - LHVM) produced a two turn alpha helix against the template sequence ' LSKPIKRPKLKTI' between the

residues 1193-1205 that has three turn alpha helix (Fig 14D, 15). The three residue gap in the query sequence against 'RPK' in the template could not be modeled as a part of alpha helix instead was modeled as the continuation of the loop. One of the reasons for this could be the absence of Proline in the query that is responsible for turns in the protein sequences. The residues Q and T in the positions 120 and 121 in the query were modeled against the break in the template sequences against Q and G in positions 1214 and 1215 respectively.

#### 3.3.3.3 A 3D Model for EIN4 Receiver Domain Using 3C3M.A as Template

3C3M.A is the PDB ID for crystal structure of the N-terminal domain of response regulator receiver protein from *Methanoculleus marisnigri* JR1. The model with lowest RMSD among all the models generated against 3C3M.A was the $5^{th}$ model with an RMSD value of 0.43Å when superimposed in MOE and $2^{nd}$ model with 0.134Å and $5^{th}$ model with 0.268Å when superimposed in PyMol. The predicted secondary structure closely resembled the crystal structure 1DCF.A, receiver domain of ETR1, with slight variations in their lengths and orientations, consisting of a beta strand at the N terminus, followed by five alpha helices and four more beta strands, all alternating with each other. The components of secondary structure were connected by loops and turns. The orientation of the fifth alpha helix towards the C-terminal end of the protein is oriented at an angle of 150° to that of 1DCF.A. Also, the second alpha helix only had two turns compared to the three turned alpha helix of 1DCF.A. The sequence identity and similarity of the query sequence with the template were 25.0 and 37.9 respectively and were aligned using BLOSUM35.

Figure 16A shows significantly diverged regions and possible orientations of the diverged regions of different models generated against 3C3M.A and superimposed the same with the template crystal structure. Structurally diverged regions were in the loop connecting second

alpha helix to the third beta strand spanning the sequence 'SNVEMSYR' in the query from residues 40 to 47 in the 5$^{th}$ model with least RMSD value. This represents the amino acids from residue 43 - 48 in the template with the sequence 'NATP - - PD'. The model generated by MOE had a short single turn alpha helix against the loop in the template, represented by a gap in the linear sequence (Fig 16C, 17).

A slight variation was observed at the alpha helix which can be considered negligible, at residue position 70-72 in the query with amino acid sequence 'FCG' (Fig 16D, 17).

#### 3.3.3.4 Comparison of the Three Predicted Structures of EIN4

The models generated using the templates with lowest RMSD values when superimposed against the respective templates, 1DCF.A**,** 2R25.B and 3C3M.A were superimposed against each other in PyMol to look at the regions that were conformationally different among the three models and the results were tabulated (Data not shown) (Fig 18).

The structures significantly have the loop regions diverged from others all throughout. The basic blue print of the secondary structure remained similar for most of the protein, with some exceptions in the lengths of the helices or beta strands and orientations of the backbone carbons. The loop connecting the second alpha helix and the beta strand was conformationally very differet in all the proteins, as it created short alpha helices in the models generated against1DCF.A and 3C3M.A (Fig 18).

It was interesting to note that the γ-loop which is involved in various key physiological functions of the receptor was oriented similarly in models generated against 2R25.B and 3C3M.A but is at an angle of about 150° in case of that generated by 1DCF.A. The third alpha helix looked much different in each of the models and the beta helix following it was the shortest in 2R25.B. Towards the end of the protein models, the structures were very different in their length and

orientation of the end of the fifth alpha helix and a stretch of tail sequence. The alpha helices of 1DCF.A and 2R25.B were similar with different orientations of their tail but the alpha helix as well as the C terminal tail for 3C3M.A were oriented at about 150° to the other two (Fig 18).

## 3.4 Comparison of ETR2 and EIN4 Models with ETR1 Receiver Domain Crystal Structure

The analysis of all the models generated for ETR2 and EIN4 receiver domains suggested that the fourth model generated has the lowest RMSD value when superimposed against the crystal structure of ETR1 receiver domain, 1DCF. They were superimposed on PyMol and the diverged regions were analyzed. The blue print of the secondary structure in both the models was similar to that of 1DCF, five beta strands and five alpha helices alternating with each other, varying in lengths. The loops connecting these were diverged as most of them represent gaps in the linear sequence of either template or query sequence, that could not be modeled accurately. The RMSD values were tabulated (Data not shown).

The second alpha helix in the crystal structure had two turns while another extra turn was generated in the model generated for ETR2 receiver domain. The model generated for EIN4 receiver domain generated a short half turn alpha helix in the loop connecting the second alpha helix to the third beta strands, which by itself was conformationally very diffrent in structure. The γ-loops were oriented in the same direction in both the models and were structurally aligned well. However, the models generated using other templates suggested that the γ-loop might be in the opposite orientation to that of 1DCF.A. The third alpha helix was shorter in both the models with about two and a half turns. The loop connecting it to the fourth beta strand was similar to 1DCF.A in case of EIN4 but was shorter and was oriented differently in case of ETR1. The C-

terminal end was oriented well with 1DCF.A for EIN4 but was different in ETR2, with a shorter alpha helix (Fig 19).

## 3.5 Conclusions

The physiology experiments suggested that ETR1 receiver domain has little or no role in mediating responses to ethylene in the presence of silver. The models generated against 1DCF.A had the lowest RMSD values suggesting that the receiver domains of ETR2 and EIN4 are closer both structurally and sequence wise, compared to any other template suggested for each of these receptors. Superimposing the final models of ETR2 and EIN4 with 1DCF. A suggested that the secondary structure is very similar to that of ETR1 receiver domain and the only difference being the loops that are generated to compensate gaps during alignment. The γ-loops are oriented in the same direction in both the models and are structurally aligned well. However, the models generated using other templates suggested that the γ-loop might be in the opposite orientation to that of 1DCF.A. This might give interesting insights into the varied modulation of the physiological responses by the receiver domain.

# 4 HOMOLOGY MODELING OF THE KINASE DOMAINS

## 4.1 Introduction

The structure of proteins involved in signal transduction in ethylene receptors is similar to that of the bacterial two component signaling, consisting of a kinase domain and an additional receiver domain (only in ETR1, ETR2 and EIN4) (O'Malley and Bleecker., 2003). The kinase domain is composed of a two main components: One is a dimerization and histidine phosophotransfer sub domain (DHp), which consists of a characteristic H box. The DHp domain is required for histidine protein kinase homo dimer formation, that is necessary for auto phosphorylation through H box histidine (Wolanin *et al.,* 2002). The second component is a catalytic transmitter domain and consists of N, G1, F and G2 box. Each of these sub domains are mainly involved in binding ATP. The N box coordinates divalent metal ions required for ATP binding, and the F, G1 and G2 boxes coordinate together to facilitate ATP lid function (Parkinson and Kofoid., 1992; Stewart., 2010; Voet-van-Vormizeele and Groth., 2008). The subfamily I ethylene receptors consist of histidine kinases and the subfamily II have the Ser-Thr kinase activity. However, ERS1 of subfamily I acts as both Histidine and Ser-Thr kinase (Moussatche., 2004).

In ETR1, signal perception by the amino-terminal domain controls the auto phosphorylation of a conserved histidine residue in the catalytic domain. The phosphoryl group is then transferred to a conserved aspartate in receiver domain. The phosphate is transferred to the histidine containing phosphotransfer protein which activates various responses to ethylene (Grefen and Harter, 2004). However, there is data suggesting this biochemical activity does not lead to ethylene responses. Binding of ethylene was shown to reduce autophosphorylation of ETR1 in the presence of copper as a metal cofactor (Voet-van-Vormizeele and Groth, 2008). In the presence of ethylene

antagonists such as silver ions or 1-Methyl cyclopropene (1-MCP), the reduction in ETR1 phosphorylation was inhibited (Voet-van-Vormizeele and Groth, 2008). The kinase domain was shown to be required for modulation various physiological processes in *Arabidopsis*. It was shown to modulate growth recovery after ethylene removal in dark grown *Arabidopsis* seedlings (Binder *et al.,* 2004b). Whether the kinase domain is involved in other physiological and biochemical aspects in plants is yet to be extensively studied.

## 4.2 Results

There is not much information available regarding the crystal structures of ethylene receptors of *Arabidopsis* in the protein data bank, except that a recent study by Mueller-Dieckmann group determined the crystal structure of catalytic ATP binding domain and predicted the structure of the entire cytosolic domain of ETR1 using previous data on the crystal structure of receiver the domain, SAXS data for the kinase domain (PDB ID:4PLA.A) and homology modeling of the GAF domain (Mayerhofer *et al.,* 2015). This chapter mainly focuses on predicting three dimensional structures of kinase domains in all the five receptors in *Arabidopsis*.

The basic structure of 4PL9.A consists of an N-terminus end starting with a stretch of residues followed by the first beta strand, followed by an alpha helix with 4-5 turns, and then two beta strands separated by loop and turn, followed by another alpha helix with 4-5 turns, then comes another beta strand. The structure shows a break and an immediate beta strand followed by a shorter alpha helix and another helix with 3-4 turns followed by two beta strands towards the C-terminus. Another break was shown in the structure between the two alpha helices before the C-terminal beta strands.

### 4.2.1 Sequence Alignment and Similarity Between Kinase Domains of Ethylene Receptors in Arabidopsis.

The sequences that are evolutionarily related have structural similarities. We used CLUSTAL W2 (an improved version of CLUSTAL W), a multiple sequence alignment tool for divergent proteins (Thompson *et al*., 1994), to compare the sequence identity and similarity between the kinase domains of ethylene receptors (Fig 20A,B). The residues that are indicated with an asterisk (*) denote identical residues in all sequences that are conserved, colons (**:**) denote highly conserved substitutions and periods (**.**) denote semi conserved substitutions indicating weakly conserved residues (Chenna *et al.*,2003). The receptors belonging to each subfamily had higher sequence similarity within themselves compared to those of the other subfamily (Fig 20B).

### 4.2.2 Predicting Three Dimensional Structure of ETR1 Kinase Domain using Homology Modeling

ETR1 belongs to the subfamily I of the ethylene receptors in *Arabidopsis* and its kinase domain spans the residues from 350-585. A list of target templates were selected using PDB from MOE, based on Z value. The query was modeled against ten templates that were chosen by MOE based on highest Z scores (Table 3). Most of them were sensor histidine kinases involved in signal transduction pathways of several bacteria and other organisms. The RMSD values were tabulated for comparative analysis. It was interesting to note that the templates with higher sequence identity/similarity/Z scores compared to others did not always yield models with commensurate lower RMSD values. To assess which was a more better model, the models were also aligned to the crystal structure in PyMol and the values were tabulated. PyMol was found to have a better alignment of the backbone carbon atoms.

Graphs were plotted with PDB ID on X-axis and RMSD values on Y-axis for both MOE and PyMol superimposed structures). The average RMSD value of the templates superimposed in MOE was around 1.107Å and that of PyMol was approximately 0.810Å. This calculation

excludes the template 4A2L. E, which had higher RMSD values (>10) in both the cases. Overall comparison of sequence identity/similarity and RMSD values suggest that models generated against the templates 2C2A.A and 3DGE.A would be approximately closer to the three dimensional structure of ETR1 kinase domain.

#### 4.2.2.1 A 3D Model of ETR1 Kinase Domain Using 2C2A.A

2C2A.A is the PDB ID for the structure of the entire cytoplasmic portion of a sensor histidine kinase protein in *Thermotoga maritima.* The model with lowest RMSD among all the models generated against 2C2A.A was the $9^{th}$ model with an value of 1.13Å when superimposed in MOE and $10^{th}$ model with 0.29Å and RMSD of $9^{th}$ model was 0.44Å when superimposed in PyMol. The predicted secondary structure closely resembled the crystal structure 4PL9.A, catalytic subunit of the kinase domain of ETR1, which constitutes only a part of the query sequence, with differences mainly in the loop orientations an lengths of beta strands. The rest of the sequence that constitutes the DHp domain was modeled very similar to that of 2C2A.A, with two extended large alpha helices connected by a loop.

The N-terminus started a little after that of the template and the stretch of residues from 16-34 in the query sequence with the amino acids 'LSS - - - - LLQETELTPEQRLMVE' modeled a long loop for parts of alpha helices and the loop connecting them in the template. This can be justified by the presence of a gap of four residues in the template against the residues TIYN in the template (Fig 21C). The second beta strand was modeled as a shorter one compared to the template at the residues, 90-96 in the query with the sequence 'LPITLNL'. The third and fourth beta strands were modeled as two short beta strands and a very long loop with different orientations extending across the next alpha helix and further. This can be associated with various gaps in both the template and the query sequences on aligning the sequences. This long

stretch extends from the residues 138-198 with the sequence 'TSKDTRAADFFVVPTGSHFYLRVKVKDSGAGINPQDIPKIFTKFAQTQSLATRSSGGSGLG' in the query sequence (Fig 22). The last two beta strands towards the end of the C-terminus were modeled as shorter ones due to in continuous gaps in the template from residues 462 - 480 with the sequence 'VESEV - GKGSR - - FFVWIPKDR'.

#### 4.2.2.2 A 3D Model for ETR1 Kinase Domain Using 3DGE.A

3DGE.A is the PDB ID for structure of a histidine kinase-response regulator complex in *Thermotoga maritima*. The model with lowest RMSD among all the models generated against 3DGE.A was the 4$^{th}$ model with an value of 0.56Å when superimposed in MOE and 0.25Å when superimposed in PyMol. The predicted secondary structure closely resembled the crystal structure 4PL9.A, catalytic subunit of the kinase domain of ETR1, which constitutes only a part of the query sequence, with differences mainly in the loop orientations an lengths of the beta strands. The rest of the sequence that constitutes the DHp domain was modeled very similar to that of 3DGE.A, with two extended large alpha helices connected by a loop.

The major conformational differences were in the loop connecting the first two alpha helices towards the N-terminus. The loop was modeled against a few residues for the first alpha helix towards its end and continues into the second alpha helix, at the residues 16-22 (LS - - - - SLLQE). The 2-3 turn in the second alpha helix was modeled as a short loop due to a valine missing at the position 293 in the template sequence, and then the structure continues as the second alpha helix, justified by the presence of gaps in the alignment (Fig 23C). The second beta strand was modeled shorter than that of the template at the position 90-97 in the query with sequence 'LPITLNLA'. The fourth beta strand and the loop connecting it to the immediate beta

strand was also modeled as two short beta strands at the residues 135-146 'ALVTKSDTRAAD' in the query. The serine at position 140 in the template had a gap in the template sequence.

The fifth beta strand was modeled as two very short beta strands connected by a hair pin turn in the query sequence (Fig 23D). This part of the query sequence corresponds to a large gap in the template from residues 407 - 411 represented by 'II - - - - - - - - - - VED' and continued as a loop connecting the rest of the protein (Fig 24). A gap of one residue in the template between 464 and 465 (S - E) was against an aspartate in the query sequence and generated a loop in the place of the end of the beta strand in the template. (Fig 24)

### 4.2.3 Predicting Three Dimensional Structure of ERS1 Kinase Domain Using Homology Modeling

ERS1 belongs to the subfamily I ethylene receptors in *Arabidopsis* and its kinase domain spans the residues from 350-589. A list of target templates were selected using PDB from MOE, based on Z value. The query was modeled against ten templates that were chosen by MOE based on highest Z scores (Table 4). Most of them were sensor histidine kinases involved in signal transduction pathways of several bacteria and other organisms. The RMSD values were tabulated for comparative analysis. It was interesting to note that the templates with higher sequence identity/similarity/Z scores compared to others did not always yield models with commensurate lower RMSD values. To assess the choice of better and accurate model, the models were also aligned to the crystal structure in PyMol and the values were tabulated. PyMol was found to have a better alignment of the backbone carbon atoms.

Graphs were plotted with PDB ID on X-axis and RMSD values on Y-axis for both MOE and PyMol superimposed structures. The average RMSD value of the templates superimposed in MOE was around 1.458Å and that of PyMol was approximately 0.627Å. This calculation excludes the template 4JAS.A, which had higher RMSD values (>10) in both the cases. Overall

comparison of sequence identity/similarity and RMSD values suggest that models generated against the templates 2C2A.A and 3DGE.A would be approximately closer to the three dimensional structure of ERS1 kinase domain. 3SL2.A seemed to be a competent template but it had lower sequence identity and similarity values compared to the other two templates. 4PL9.A was not used as a template to predict the structure because the DHp subunit of kinase domain was not modeled using 4PL9.A

#### 4.2.3.1 A 3D Model for ERS1 Kinase Domain Using 2C2A.A

2C2A.A is the PDB ID for the structure of the entire cytoplasmic portion of a sensor histidine kinase protein in *Thermotoga maritima*. The model with lowest RMSD among all the models generated against 2C2A.A was the 8$^{th}$ model with an value of 0.66Å when superimposed in MOE and 3$^{rd}$ model with 0Å when superimposed in PyMol. Hence, model 3 was used as the best model. The secondary structure looked the same compared to 4PL9.A for the catalytic domain but was oriented differently with extra beta strands and turns and loops connecting them.

The DHp subunit consisted of two major alpha helices and the a part of the first alpha helix and the loop connecting it to the second one was mainly modeled as a long loop, due to a gap present in the query sequence from 13 - 26 represented by 'II - - - - SLSSLLLETELS' (Fig 25C). The loop connecting the fourth alpha helix and the fourth beta strand was found to be highly variable as it showed two beta strands connected by a turn. This part of the query with sequence 'KEGYISIIASIMKPESLQEL' from 127-146 was modeled against a large gap in the template sequence from 385 - 389 (KK - - - - - - - - - - - - - - - - DA) (Fig 25D, 26).

There was a large loop formation between the fifth short alpha helix and sixth alpha helix, which corresponded to a break in the template structure in which the ends of the breaks were oriented in different angles, from the residues 186 - 205 in the query (VQPRTGTQRNHSGGGLGLAL).

The loop connecting the last two beta strands was slightly diverged as there was a gap in the template sequence from 465 - 468 with sequence 'E - - - VGK' against 'GLEKGC' in the query sequence from 223 - 228 (Fig 25C, 26).

#### 4.2.3.2 A 3D Model for ERS1 Kinase Domain Using 3DGE.A

3DGE.A is the PDB ID for structure of a histidine kinase-response regulator complex in *Thermotoga maritima*. The model with lowest RMSD among all the models generated against 3DGE. A was the $10^{th}$ model with an value of 0.46Å when superimposed in MOE and 0.172Å when superimposed in PyMol. The DHp domain was modeled very similar to that of 3DGE.A, with two extended large alpha helices connected by a loop, but were closer in orientation to the catalytic domain compared to 2C2A.A.

Some regions in the first two alpha helices towards the N-terminus were modeled with some loops due to gaps in the query from 13 -18 with sequence 'I - - - - ISLSS' and another short loop in the second alpha helix from 30 -32 'RVM' due to a gap in the template sequence (E - F) (Fig 27C, 28). The second beta strand was modeled as two short strands connected by a loop across the length of the same in the template, from 91-97 in the target sequence 'STNLILS' . The loop connecting the fourth alpha helix and the fourth beta strand was oriented in a different direction against a large gap in the template sequence from 387 - 379 (K - - - - - - - - - - - - - - - DA) (Fig 27 D, 28). The end of the fourth beta strand was extended into another short beta strand with sequence 'LSDS' from residues 155-158, which was a part of the loop in the template. The loop connecting the last two alpha helices was slightly conformationally different in the regions 'EGLE' from 222 -225as there was a gap in the template (E - V) at position 465 and 466.

### 4.2.4 Predicting Three Dimensional Structure of ETR2 Kinase Domain Using Homology Modeling

ETR2 belongs to the subfamily II ethylene receptors in *Arabidopsis* and its kinase domain spans the residues from 374-614. A list of target templates were selected using PDB from MOE, based on Z value. The query was modeled against ten templates that were chosen by MOE based on highest Z scores (Table 5). Most of them were sensor histidine kinases involved in signal transduction pathways of several bacteria and other organisms. The RMSD values were tabulated for comparative analysis. To assess the choice of better and accurate model, the models were also aligned to the crystal structure in PyMol and the values were tabulated. PyMol was found to have a better alignment of the backbone carbon atoms (Table 4).

Graphs were plotted with PDB ID on X-axis and RMSD values on Y-axis for both MOE and PyMol superimposed structures. The average RMSD value of the templates superimposed in MOE was around 1.27Å and that of PyMol was approximately 0.497Å. This calculation excludes the templates 3LQ3.A, 3WIQ.A, 4CMP.A, 4MYJ.A, 4OO8.A, 3FEG.A, and 4UN3.A which had higher RMSD values (>10) in both the cases. Overall comparison of sequence identity/similarity and RMSD values suggested that models generated against the templates 2C2A.A and 3A0Y.B (although it had lower sequence identity and similarity with the target compared to their templates) would be approximately closer to the three dimensional structure of ETR2 kinase domain. 4PL9.A was not used as a template to predict the structure because the DHp subunit of kinase domain was not modeled using 4PL9.A. It was interesting to note that the templates with higher sequence identity/similarity/z scores compared to others did not always yield models with commensurate lower RMSD values. Although other template had a higher sequence alignment/similarity with the target sequence, the model generated against the showed very high RMSD values.

#### 4.2.4.1 A 3D Model for ETR2 Kinase Domain Using 2C2A.A

2C2A.A is the PDB ID for the structure of the entire cytoplasmic portion of a sensor histidine kinase protein in *Thermotoga maritima.* The model with lowest RMSD among all the models generated against 2C2A.A was the 7$^{th}$ model with an value of 0.81Å when superimposed in MOE and 7$^{th}$ model with 0.382Å when superimposed in PyMol. Hence, model 7 was considered as the optimal model.

The DHp subunit consisted of two major alpha helices and the end of the first helix and the loop connecting it to the second helix had a slight divergence in the region spanning the residues 18-23 with the sequence 'I - - - QDEK' justified by the presence of gap in the target sequence. Another gap in the target sequence from 51 - 53 (D - - - VP) generated a loop towards the end of second alpha helix. The loop connecting the fourth and fifth beta strands had a major gap in the template which showed beta strands in different orientation from the template loop, connected by a hair pin turn, followed by a short beta strand and another loop against the length of fifth beta strand with residues from 137-173, 'RGSLDRSDHRWAAWRSPASSADGDVYIRFEMNVEND'. (Fig 29)

The fifth alpha helix was partially modeled as the target model had the loop extended into the first turn of this short helix for the residues 179 - 185 (SFASVSS). The break in the loop connecting the fifth and sixth alpha helices was modeled into a larger loop with a different orientation than that of the ends of the break, followed by a short beta strand. This region spanned the residues 189 - 207 in the query sequence 'EVGDVRFSGGYGLGQDLSF' which was against the gaps and break in the template sequence. The loop connecting the last two beta strands had a gap in the query sequence, and hence a slight divergence in the length of the loop 227 - 229 (S - DG) (Fig 30).

#### 4.2.4.2 A 3D Model for ETR2 Kinase Domain Using 3A0Y.B

3A0Y.B is the PDB ID for Catalytic domain of histidine kinase ThkA in *Thermotoga maritima.* The model with lowest RMSD among all the models generated against 3A0Y.B was the 10$^{th}$ model with an value of 0.83Å when superimposed in MOE and 0.512Å when superimposed in PyMol.

The model did not seem have the DHp domain as the template constitutes mainly the catalytic domain. The loop between the first and the second beta strand diverged slightly due to a gap in the template 'E - D' at position 640 and 641, that corresponds to the sequence 'SLP' at 95-97 in the target sequence. A part of second alpha helix and the loop connecting it to the third beta strand was conformationally very different with a different orientation of the loop to cover the gap in the template from 664 - 671 (E - - - - ATGENGK). The successive beta strand ended as a short one in the model, followed by a diverged loop in the template region 678 - 682 with 'D - - - MYTK'. (Fig 31)

The loop connecting the fourth beta strand and the third alpha helix was structurally diverge to fill a gap in the template from 694 - 699 spanning from 'IPE - - - - - - - - - - - ELK'. The fourth alpha helix was kinked and the model had a long loop in this region, which in turn consisted of a very short alpha helix and an extended loop into half of the length of the alpha helix in the template, to accommodate the gap from residues 712 - 719 with sequence 'QG - - - - - - - - - - - - TGLG - - - - LS' (Fig 32).

### 4.2.5 Predicting Three Dimensional Structure of ERS2 Kinase Domain Using Homology Modeling

ERS2 belongs to the subfamily II ethylene receptors in *Arabidopsis* and its kinase domain spans the residues from 389-623. A list of target templates were selected using PDB from MOE, based on Z value. The query was modeled against ten templates that were chosen by MOE based on

highest Z scores (Table 6). The RMSD values were tabulated for comparative analysis. To assess the choice of better and accurate model, the models were also aligned to the crystal structure in PyMol and the values were tabulated. PyMol was found to have a better alignment of the backbone carbon atoms.

Graphs were plotted with PDB ID on X-axis and RMSD values on Y-axis for both MOE and PyMol superimposed structures. The graphs excluded the templates 2IV7.A, 2IW1. A, 3BGA.A, 4JN5. A, 4DIQ.A, and 2BW3. A which had higher RMSD values (>10) in both the cases. Overall comparison of sequence identity/similarity and RMSD values suggest that models generated against the template 3CVR.A would be approximately closer to the three dimensional structure of ERS2 kinase domain. But, it is a ligase and is ideally comparable to a kinase protein. Hence 4PLA.A was chosen as the template. The limitation to this analysis would be that the DHp domain of ERS2 kinase domain cannot be visualized.

#### 4.2.5.1 A 3D Model for ERS2 Kinase Domain Using 4PL9.A

4PL9.A is the PDB ID for the structure of the catalytic domain of ETR1 from *Arabidopsis thaliana.* The model with lowest RMSD among all the models generated against 4PL9.A was the 1$^{st}$ model with an value of 0.65Å when superimposed in MOE and with 0.229Å when superimposed in PyMol. Hence, model 1 was chosen as the optimal model.

The loop connecting third and fourth beta strands consisted of a break in the template crystal structure and the model generated in this region consists of a large loop with a turn, to accommodate the gaps between the residues in both the target and template sequences (Fig 23 C,24) from residues 136 - 157 in the query sequence (PESGNSDVSERKDIQEA - - - - - - - - - - VWRHC). Another large loop was formed in the break between the third and fourth alpha helices, from residue 183 - 197 with sequence 'S - - - - - - - - - - - GSNLEEEEENPSLN'. The

loop connecting the last two beta strands was larger than that of the template, with more residues in the query sequence to fill in the gap, between 570 - 572 (G - - KG). (Fig 33, 34)

### 4.2.6 Predicting Three Dimensional Structure of EIN4 Kinase Domain Using Homology Modeling

EIN4 belongs to the subfamily II ethylene receptors in *Arabidopsis* and its kinase domain spans the residues from 374-612. A list of target templates were selected using PDB from MOE, based on Z value. The query was modeled against ten templates that were chosen by MOE based on highest Z scores (Table 7). To assess the choice of better and accurate model, the models were also aligned to the crystal structure in PyMol and the values were tabulated. PyMol was found to have a better alignment of the backbone carbons.

Graphs were plotted with PDB ID on X-axis and RMSD values on Y-axis for both MOE and PyMol superimposed structures. The graphs excluded 1OBH.A, 2V0C.A, and 3PIH.A which had higher RMSD values (>10) in both the cases. Overall comparison of sequence identity/similarity and RMSD values suggested that models generated against the templates 2C2A.A and 4Q20. A(superseded version of 4EW8.A) would be approximately closer to the three dimensional structure of ETR2 kinase domain. 4PL9.A was not used as a template to predict the structure because the DHp subunit of kinase domain was not modeled using 4PL9.A.

#### 4.2.6.1 A 3D Model for EIN4 Kinase Domain Using 2C2A.A

2C2A.A is the PDB ID for the structure of the entire cytoplasmic portion of a sensor histidine kinase protein in *Thermotoga maritima.* The model with lowest RMSD among all the models generated against 2C2A.A was the $4^{th}$ model with an value of 1.12Å when superimposed in MOE and $4^{th}$ model with 0.372Å when superimposed in PyMol. Hence, model 4was chosen to be the optimal model.

The DHp subunit consisted of two major alpha helices and the end of the first helix had a slight divergence in the region spanning the residues 16-23 with the sequence 'LSS - - - -LLQET' justified by the presence of gap in the target sequence. The second alpha helix had a small loop around the second turn due to a gap in the template sequence from 292 - 293 with the residues ;L - E'. (Fig 35)

The fifth and sixth beta strands and the loop connecting them consisted of two beta short beta strands corresponding to those of the template and a large loop with two turns connecting them around the residues 405 - 409 in the template with the sequence 'LI - - - - - - - - - - - IV'. A long stretch of sequence produced a long loop at the break between the fifth and sixth alpha helices, compensating various gaps in both query and template sequences between 170 - 202 and 417 - 449 respectively. The loop seemed to have been extending into the second alpha helix in the DHp domain. The last two beta strands towards the C-terminus were modeled as shorter ones with a long loop connecting them, due to gaps in template sequence 462 - 473 (VESEV - GKGS - - RFF'. (Fig 36)

#### 4.2.6.2 A 3D Model for ETR2 Kinase Domain Using 4Q20. A

4Q20.A is the PDB ID for Crystal structure of a C-terminal part of tyrosine kinase (DivL) from *Caulobacter crescentus*. The model with lowest RMSD among all the models generated against 3A0Y.B was the 3$^{rd}$ model with an value of 0.80Å when superimposed in MOE and 0.306Å when superimposed in PyMol.

The loop at the end of the first alpha helix in the model started a little earlier with a slight divergence in the orientation of the beginning of the loop due to a missing residue at the position 20 - 21 'F - Q' against 'LER' at 566 - 568 in the template. The template structure had a short alpha helix between the first and second beta strands which in the model was represented as a

continuous loop between the two beta strands at the positions, 96 - 103 (VQTRLPNL) against 'CEEDV - GL' in the template from 643 - 648. (Fig 37)

The third beta strand was modeled as a large loop oriented about 180° to the template sheet, consisting of two beta strands connected by a turn, to compensate a gap in the template from 680 - 686, with the sequence 'LS - - - - - - - - - - - - ARRAL' The fifth alpha helix was also modeled as a long loop oriented perpendicularly to the template with two beta strands connected by a turn, against the large gap in the template sequence at residues 705 - 710 with the sequence 'QA - - - - - - - - - - - - - - HIFD'. The last two beta strands towards the C-terminus were entirely modeled as loops with a couple of turns, against the template sequence 'WVALESEPGNGSTFTCHLPETQ' in the position 737 - 758. (Fig 38)

## 4.3 Comparison of Models Generated Against 2C2A.A in all the Ethylene Receptors

The structure of the sensor histidine kinase, 2C2A.A was suggested by MOE as one of the potential templates for all the ethylene receptors except ERS2. Hence, we superimposed models generated for these receptors, that were not very conformationally different using 2C2A.A as template in MOE and looked at the conserved / similar and diverge regions among the receptors (Fig 39 A and B).

The first two alpha helices that constitute the DHp region of kinase domain were found to be common in all the receptors with a variable region at the end of the first helix and the loop connecting it to the second alpha helix. This was justified by the presence of a gap in all the receptor isoforms when superimposed with the template. The first part of CA domain looks very similar and the residues were conserved for the first beta strands of the subunit followed by an alpha helix and a turn, two more beta strands and another alpha helix.

The regions that were conformationally very different started around residues 120 - 125 and continued to generate gaps and hence differently oriented loops and an unusual pair of beta strands connected by a turn in ETR2 and until almost the end of the protein but especially until the residues 185 - 195 in the receptors. (Fig 40)

## 4.4 Conclusions

The templates selected by MOE for homology modeling were mostly histidine kinases and other types of sensor kinases including tyrosine kinases, in various other organisms, especially micro organisms. The templates for ETR1 and ERS1 kinase domains were both sensor histidine kinase proteins illustrating the relationship between histidine kinases in bacteria to that of ETR1 and the hybrid histidine kinase and serine-threonine kinase of ERS1. The catalytic domain of the predicted structures closely resemble that of the crystallized ETR1 CA sub unit (PDB ID: 4PL9.A). The DHp domain in both the models consisted of two large alpha helices connected by a turn that extends it the catalytic subunit.

The sensor histidine kinase 2C2A.A was one of the potential templates for kinase domain of ETR2 and EIN4, that belong to the subfamily II of ethylene receptors. This suggested that ERS2 kinase domain is structurally diversified from all the other four receptors. The other template for ETR2 was 3A0Y.B which mainly constituted the catalytic domain of histidine kinase. Hence, in this part of modeling, using 3A0Y.B as the template, the DHp domain was not very accurately modeled. For ERS2, the only apt template, we could arrive at was 4LP9.A and hence he DHp domain was not accurately modeled for ERS2.

Apart from 2C2A.A, the kinase domain of EIN4 was predicted using 4Q20.A which was tyrosine kinase in *Caulobacter crescentus*. The models generated using these receptors had DHp domain

that was similar to that of sub family I receptors and the catalytic domain similar to 4PL9.A with differences in the lengths of the alpha helix or beta strands and orientation of their loops. Comparison of models generated against 2C2A.A to all receptors except ERS2 suggested that the DHp region and other alpha helices in the CA subunit of the kinase domain is conserved among all those receptors and the regions that are conformationally different lie within the beta strands, especially in the orientation of the loops connecting them and the length of the beta strands.

# 5 HOMOLOGY MODELING OF GAF DOMAIN OF ETHYLENE RECEPTORS

## 5.1 Introduction

The GAF domain of ethylene receptors in *Arabidopsis* is located immediately after the ethylene binding domain (transmembrane domain) and before the kinase domain. GAF domains were initially identified in cGMP-specific and -stimulated phosphodiesterases, Adenylatecyclases and the *Escherichia coli* protein FhlA. GAF domains were generally thought to play an important role in binding of diverse molecules. GAF domains in cyanobacteria (*Synechocystis* sps.) has been shown to bind phycocyanobilin, a chromophore and thus involved inregulation of light (Ulijasz*et al.,* 2009). In *Arabidopsis* they are not involved in ligand binding due to a missing cysteine residue that is required for chromophore binding (Aravind and Ponting., 1997). The relevance of the GAF domains and their role in *Arabidopsis* ethylene receptor signaling and function is yet to be determined.

GAF domains are very diverged and present in many organisms, mainly involved in receptor function and mediating protein-protein interactions. Many GAF domains are being crystallized and the protein data bank is often being updated with new structures. But, the ethylene receptor GAF domains have not been crystallized and knowledge regarding the structure of GAF domains in these receptors is limited. A recent study by the Mueller-Dieckmann group predicted the structure of GAF domains using homology modeling by MODELER (Mayerhofer*et al.,* 2015). This chapter mainly focuses on predicting three dimensional structures of GAF domains in all five receptors in *Arabidopsis*.

## 5.2 Results

### 5.2.1 Sequence Alignment and Similarity Between GAF Domains of Ethylene Receptorsin *Arabidopsis*.

Proteins that are structurally similar have evolutionary relationships and have similar sequences and amino acid make-up. We used CLUSTAL W2 (an improved version of CLUSTAL W), a multiple sequence alignment tool for divergent proteins (Thompson *et al*., 1994), to compare the sequence identity and similarity between the GAF domains of ethylene receptors (Figure 41 A,B). The residues that are indicated with an asterisk (*) denote identical residues in all sequences that are conserved, colons (**:**) denote highly conserved substitutions and periods (**.**) denote semi conserved substitutions indicating weakly conserved residues (Chenna *et al.*, 2003). The receptors belonging to each subfamily had higher sequence similarity within themselves compared to those of the other subfamily (Fig 41B).

### 5.2.2 Predicting Three Dimensional Structure of ETR1 GAF Domain Using Homology Modeling

ETR1 belongs to the subfamily I ethylene receptors in *Arabidopsis* and its GAF domain spans the residues from 158-307. A list of target templates were selected using PDB from MOE, based on Z value. The query was modeled against ten templates that were chosen by MOE based on highest Z scores (Table 8). The RMSD values were tabulated for comparative analysis. To assess which was a more accurate model, the models were also aligned to the crystal structure in PyMol and the values were tabulated. PyMol was chosen as it was found to have a better alignment of the backbone carbon atoms.

The average RMSD value of the templates superimposed in MOE was around 1.9Å and that of PyMol was approximately 4.82Å. These calculations excluded the templates from the list which had higher RMSD values (>10) in both the cases. Overall comparison of sequence identity/similarity and RMSD values suggest that models generated against the templates 1UI6.A

and 1B6A.A would be approximately closer to the three dimensional structure of ETR1 GAF domain.

#### 5.2.2.1 A 3D Model of the ETR1 GAF Domain Using 1UI6. A

1UI6.A is the PDB ID for the crystal structure of gamma-butyrolactone receptor (ArpA-like protein) in *Streptomyces coelicolor*A3 (2). The model with lowest RMSD among all the models generated against 1UI6.A was the 3$^{rd}$ model with an value of 1.21Å when superimposed in MOE and 9$^{th}$ model with 5.91Å and RMSD of 3$^{rd}$ model was 6.374Å when superimposed in PyMol.

The model typically started in the middle of an alpha helix of the crystal structure of the template and consisted of several alpha helices connected by various loops and turns. It was interesting to note that there were no beta strands in the entire template structure as well as in the model. The third alpha helix was modeled only into one turn and continued as the loop connecting it to the fourth alpha helix, spanning the sequence 'W - - M - - - - - PT' from residues 25-28 in the target sequence. About half each of the fifth and sixth alpha helices were modeled as long loop connecting both of them, at residues 60 - 84 in the query sequence, 'V - FGTSRAVKISPNSPVARLRPVSGK' to compensate gaps in the corresponding template sequence. The loop connecting the sixth and seventh alpha helix was conformationally different, modeled a short alpha helix for the seventh and immediately looped into eighth helix. The template structure had a break in the structure which was created as a loop at the residues 123 - 127, which corresponded to a large gap in the query sequence. The loop extended in to the last alpha helix towards the C-terminus was a little longer, as the target sequence had gaps from 139 - 144 with the sequence 'VEV - - VAD' (Fig 42, 43).

#### 5.2.2.2 A 3D Model of the ETR1 GAF Domain Using 1B6A.A

1B6A.A is the PDB ID for the crystal structure of human methionine aminopeptidase 2 complexed with tnp-470 in *Homo sapiens*. The model with lowest RMSD among all the models generated against 1B6A.A was the 2$^{nd}$ model with an value of 1.7Å when superimposed in MOE and had extremely high RMSD values when superimposed in PyMol, for which the reason was unknown.

The template was a complex human protein that consisted of many alpha helices and beta strands. But, the model could only replicate some four alpha helices at corresponding residues to the template structure and the rest of the model consisted of long loops and turns. The similarities were in the linear sequence (Fig 44, 45). The two beta strands modeled were very short.

### 5.2.3 Predicting Three Dimensional Structure of ERS1 GAF Domain Using Homology Modeling

ERS1 belongs to the subfamily I ethylene receptors in *Arabidopsis* and its GAF domain spans the residues from 158-307. A list of target templates were selected using PDB from MOE, based on Z value. The query was modeled against ten templates that were chosen by MOE based on highest Z scores (Table 9). The RMSD values were tabulated for comparative analysis. To assess which was a more accurate model, the models were also aligned to the crystal structure in PyMol and the values were tabulated. PyMol was chosen as it was found to have a better alignment of the backbone carbon atoms. Overall comparison of sequence identity/similarity and RMSD values suggest that models generated against the templates 3TTG.A and 3FY4.A would be approximately closer to the three dimensional structure of ERS1 GAF domain.

#### 5.2.3.1 A 3D Model of the ERS1 GAF Domain Using 3TTG.A

3TTG.A is the PDB ID for crystal structure of putative aminomethyl transferase from *Leptospirillum rubarum*. The model with lowest RMSD among all the models generated against 3TTG.A was the 2$^{nd}$ model with a value of 3.56Å when superimposed in MOE and 0.689Å when superimposed in PyMol.

The model started with an alpha helix after few residues and a little deviation in the alpha helix due to a gap present in the query at positions 15 - 17 with sequence 'K - TL'. The beta strand followed this alpha helix and the model had a short cut loop into the remaining secondary structure bypassing a beta strand and an alpha helix in the template, which is justified by gaps in the query sequence. The model is conformationally very different from the secondary structure of the template and which explained by the various gaps present in both the template and target sequences (Fig 46, 47)

#### 5.2.3.2 A 3D Model of the ERS1 GAF Domain Using 3FY4.A

3FY4. A is the PDB ID for crystal structure for photolyase in *Arabidopsis thaliana*. The model with lowest RMSD among all the models generated against 3FY4.A was the 9$^{th}$ model with an value of 1.86Å when superimposed in MOE and showed very high RMSD value when superimposed in PyMol and the reason for this was unknown.

The crystal structure started with an alpha helix and several alpha helices and two beta strands that were connected by long loops which were oriented in different directions. The model only generated alpha helices and loops in different orientations but could not generate beta strands instead had long loops along the length of the beta strands in the crystal structure.

### 5.2.4 Predicting Three Dimensional Structure of ETR2 GAF Domain Using Homology Modeling

ETR2 belongs to the subfamily II ethylene receptors in *Arabidopsis* and its GAF domain spans the residues from 187-331. A list of target templates were selected using PDB from MOE, based on Z value. The query was modeled against ten templates that were chosen by MOE based on highest Z scores (Table 10). The RMSD values were tabulated for comparative analysis. To assess which was a more accurate model, the models were also aligned to the crystal structure in PyMol and the values were tabulated. PyMol was chosen to have a better alignment of the backbone carbon atom. Overall comparison of sequence identity/similarity and RMSD values suggest that models generated against the templates 2K2N.A and 4HL7.A would be approximately closer to the three dimensional structure of ETR2 GAF domain.

#### 5.2.4.1 A 3D Model of the ETR2 GAF Domain Using 2K2N.A

2K2N.A is the PDB ID for the solution structure of a cyanobacterial phytochrome GAF domain in the red light-absorbing ground state in *Synechococcus* sps. The model with lowest RMSD among all the models generated against 2K2N.A was the $4^{th}$ model with an value of 1.48Å when superimposed in MOE and 1.026Å when superimposed in PyMol. The residues with structural divergences are marked and the N-terminus and C-terminus are indicated.

The alpha helices were aligned well in both the template ad the target structures through the entire length but the beta starnds were short and the loops connecting the secondary structures were oriented in different directions with different turns. (Fig 50, 51).

#### 5.2.4.2 A 3D Model of the ETR2 GAF Domain Using 4HL7.A

4HL7. A is the PDB ID for the crystal structure of nicotinate phosphoribosyl transferase (target NYSGR-026035) from *Vibrio cholerae.* The model with lowest RMSD among all the models

generated against 4HL7.A was the 11$^{th}$ model with an value of 20.52Å when superimposed in MOE and had 0.47Å when superimposed in PyMol.

The alignment in MOE resulted in a very high RMSD value, but the secondary structure of the model seemed to have just oriented in a different direction to the template and this was fixed by aligning them in PyMol. Most of the structure consisted of alpha helices connected by long loops and turns, reciprocated in the model but with little orientation differences to the crystal structure. The analysis was limited and could not show the similarities in the linear sequence due to lack of proper display tools to show the aligned sequences explicitly. (Fig 52)

### 5.2.5 Predicting Three Dimensional Structure of ERS2 GAF Domain Using Homology Modeling

ERS2 belongs to the subfamily II ethylene receptors in *Arabidopsis* and its GAF domain spans the residues from 190-346. A list of target templates were selected using PDB from MOE, based on Z value. The query was modeled against ten templates that were chosen by MOE based on highest Z scores (Table 11). The RMSD values were tabulated for comparative analysis. To assess which was a more accurate model, the models were also aligned to the crystal structure in PyMol and the values were tabulated. PyMol was chosen to have a better alignment of the backbone carbon atoms. Overall comparison of sequence identity/similarity and RMSD values suggest that models generated against the templates 2K2N.A and 2K7W.A would be approximately closer to the three dimensional structure of ERS2 GAF domain.

#### 5.2.5.1 A 3D Model of the ERS2 GAF Domain Using 2K2N. A

2K2N.A is the PDB ID for the solution structure of a cyanobacterial phytochrome GAF domain in the red light-absorbing ground state in *Synechococcus* sps. The model with lowest RMSD

among all the models generated against 2K2N.A was the 10th model with an value of 2.3Å when superimposed in MOE and 8.892Å when superimposed in PyMol.

The alpha helices were aligned well across the length of them in the template and the model generated only shorter ones in place of beta strands from the template. Most of the structure seemed to have been replicated in the model except that a beta strand has been bypassed entirely (Fig 53, 54)

#### 5.2.5.2 A 3D Model of the ERS2 GAF Domain Using 2K7W.A

2K7W.A is the PDB ID for the crystal structure of BAX activation at a novel interaction site in *Homo sapiens*. The model with lowest RMSD among all the models generated against 1B6A. A was the 1st model with an value of 1.36Å when superimposed in MOE and had extremely high RMSD values when superimposed in PyMol, for which the reason was unknown.

The entire crystal structure was closely replicated in the model but with slight variations in the loop regions. There were nine alpha helices, both short and long and no beta strands. The model structure started in the middle of the first alpha helix of the template. The loop connecting the first and the second alpha helix was diverged and oriented differently due to the presence of gap in the query sequence from 10 - 21 (LVELSK - - - - - - - - TLGLKN). There was a short loop formation in the beginning of the fifth alpha helix from 110 -114 in the template sequence 'VVA - - -LF. The loop connecting the fifth and sixth alpha helix was diverged and the sixth alpha helix was modeled as a shorter one due to a gap in the template from 128 - 135 in the template with sequence 'KV - - - PEL - IRT'. The loop connecting the seventh and eighth alpha helices which were short was modeled as another short alpha helix and the eighth alpha helix was represented as a loop at the residues 122 - 132 in the query sequence 'CYAILVCVLPL'. (Fig 55, 56).

### 5.2.6 Predicting Three Dimensional Structure of EIN4 GAF Domain Using Homology Modeling

EIN4 belongs to the subfamily II ethylene receptors in *Arabidopsis* and its GAF domain spans the residues from 181-331. A list of target templates were selected using PDB from MOE, based on Z value. The query was modeled against ten templates that were chosen by MOE based on highest Z scores (Table 12). The RMSD values were tabulated for comparative analysis. To assess which was a more accurate model, the models were also aligned to the crystal structure in PyMol and the values were tabulated. PyMol was chosen as it was found to have a better alignment of the backbone carbon atoms. Overall comparison of sequence identity/similarity and RMSD values suggest that models generated against the templates 2K2N.A and 2PRR.A would be approximately closer to the three dimensional structure of EIN4 GAF domain.

#### 5.2.6.1 A 3D Model of the EIN4 GAF Domain Using 2K2N.A

2K2N.A is the PDB ID for the solution structure of a cyanobacterial phytochrome GAF domain in the red light-absorbing ground state in *Synechococcus*sps. The model with lowest RMSD among all the models generated against 2K2N.A was the 2$^{nd}$ model with an value of 1.01Å when superimposed in MOE and had very high RMSD value when superimposed in PyMol.

The structure typically consisted of four alpha helices and five beta strands connected by loops of different lengths and orientations. The first beta strand was modeled as a shorter one against the template and the rest of the beta strand modeled as a loop extending into the immediate beta strand, with a slight divergence in the structure from residues 23 - 35 in the target sequence 'AVWMP - - NENRTEMH'. (Fig 57)

The long loop connecting the second beta strand to second alpha helix was modeled with slight divergence and as a continuous sequence into the third beta strand bypassing the formation of third alpha helix at the position 50 - 63 in the target sequence 'VIPIN - DPDVV - - QVRE'. The

third beta strand and the third alpha helix were connected by a long loop, with slight divergence and the gap created in the target sequence between 70 - 75 (LR - - - - KNSV' could not model a very short alpha helix as per in the template sequence. The loop connecting the last two beta strands was much diverged to compensate a gap in the query from 115 - 122 'TPY - - - - AIMVL'. (Fig 58).

#### 5.2.6.2 A 3D Model of the EIN4 GAF Domain Using 2PRR.A

2PRR. A is the PDB ID for crystal structure of alkylhydroperoxidase AhpD corein *Ralstonia eutropha*. The model with lowest RMSD among all the models generated against 2PRR. A was the 2$^{nd}$ model with an value of 0.82Å when superimposed in MOE and 0.234Å when superimposed in PyMol, for which the reason was unknown.

The structure predicted consisted only of alpha helices and no beta strands. The first alpha helix was modeled as a shorter one towards the end of the helix, from 11 - 15 in query sequence with residues 'VEL - - - - - - SK - - ILD'. The beginning of the fifth alpha helix was modeled as a loop to fill in the gaps in the sequence 75 -77 'V - - LA' against 'CLYCV' in the template sequence. The end of sixth alpha helix in the template had a kink and the target sequence could model a loop in the position from residues 100 - 106 to compensate the gap in the template from 113 -116 'NY - - - LK'. These apart, the query sequence was modeled well using the template structure. (Fig 59, 60)

## 5.3 Comparison of Models Generated from Common Template 2K2N.A for Subfamily II receptors

The receptors in the subfamily II had 2K2N.A as a common template. The models generated against this template for each of the receptors with least RMSD value were superimposed against each other (Fig 61 A, B). The basic structure consisted of alpha helices that were well conserved

over all the three receptors and five beta strands varied in their lengths and their loops were oriented in different directions. The beta strands were shorter in length and EIN4 did not have the third alpha helix at all. The conserved regions among these receptors were marked in (Fig 62).

## 5.4 Conclusions

The models generated for different sub families of GAF domains were found to be very diverged and a sub set of three receptors from the sub family II have similar or conserved regions. The sub family I receptors ETR1 and ERS1 have the alpha helices conserved in all the predicted models. The beta strands were either not modeled at all or were represented as long loops or very short beta strands were modeled across the length of the beta sheets in the template structures. ERS1 has *Arabidopsis* photolyase enzyme as one of the possible templates (3FY4.A).

The subfamily II receptors had 2K2N.A as a common template, which is structure of a cyanobacterial phytochrome GAF domain in the red light-absorbing ground state in *Synechococcus* sps. Comparison of all the models against this template suggested that the alpha helices were observed, except for one of the alpha helices in EIN4 and the rest of the regions were closely related/conserved/ similar to each other suggesting that the GAF domains of sub family II receptors have conserved domain regions that are similar to that of GAF domain of Cyanobacteria. When the sub family I receptors were modeled using 2K2N.A as template, the RMSD values were very high (~15-20) (Data not shown) when superimposed each of the models with the crystal structure. This suggested that the sub family II receptor structure was significantly different from that of those belonging sub family I and they diverged from the GAF domains of cyanobacteria.

# 6 DISCUSSION

Ethylene is a gaseous hormone and is perceived in plants through a family of five receptors, ETR1, ERS1, ETR2, ERS2 and EIN4 in *Arabidopsis*. The ethylene receptors have evolved to be sub-functionlized by performing several over lapping and non over lapping roles in *Arabidopsis thaliana* (Shakeel *et al.,* 2013). Each of the receptor domains in turn modulate various physiological responses to ethylene. Metal ions such as copper and silver were shown to be required for ethylene binding to the receptors (Rodriguez *et al.,*1999). It is interesting to note that although silver supports ethylene binding to ETR1, silver ions are known to block responses to ethylene. This led to an understanding that silver binds to the receptor in the transmembrane ethylene binding domain but fails to allow necessary conformation in the receptor that is required for further downstream signaling. McDaniel and Binder in 2012 proved that "Ethylene receptor 1 (ETR1) is sufficient and has the predominant role in mediating inhibition of ethylene responses by silver". This led to the question about the sub-functionalization of each of the receptor domains in responses to silver.

A part of the present study identified that ETR1 receiver domain has little or no role in mediating responses to silver ions, supported by data obtained from end point analysis and analyzing growth kinetics of dark grown *Arabidopsis* seedlings. However, it is known to be important for other responses (Wilson *et al.,* 2014). This led to an interest in studying the structural aspects that lead to the sub-functionalization of ethylene receptors. This can be achieved by visualizing the structure of the cytosolic domains of the ethylene receptors in *Arabidopsis*. Unfortunately, the information regarding the crystal structure of different domains of ethylene receptors is only limited to the receiver domain of ETR1 (PDB ID: 1DCF, Muller-Dieckmann *et al.,* 1999) and catalytic ATP-binding domain of ETR1 (Mayerhoff *et al.,* 2014). To overcome this, we resorted

to predicting the three dimensional protein structure using knowledge-based prediction homology modeling.

Protein that are structurally similar have evolutionary relationships and have similar sequences and amino acid make-up. The sequence similarity among receptors in each domain suggested that the receptors that belong to each sub family are closer in sequence compared to those in the other sub family, which explains the concept of having them place in different sub families. It was interesting to note that although they were similar sequence wise, they were structurally and functionally diverse. Through this study, we could generate homology models against various templates chosen by MOE for a given query sequence. Knowledge-based prediction helped to predict the three dimensional structure of a query protein sequence. The target and the template sequence were aligned using BLOSUM matrices with least gap penalty. The structures were prepared and protonated and the models were generated using CHARMM27. Choosing the right matrix and force field were quintessential for accurate model generation.

Only a sub set of ethylene receptors contain receiver domain at the end of the C-terminus of the receptor. The models generated showed similar tertiary structure for ETR2 and EIN4 receiver domains as compared to that of ETR1 crystal structure. Loop alignments were different at certain regions that represented gaps in the sequences. The γ-loops are oriented in the same direction in both the receptor models and were structurally aligned well. However, the models generated using other templates apart from 1DCF. A suggested that the γ-loop might be in the opposite orientation to that of 1DCF.A. This might give interesting insights into the varied modulation of the physiological responses by the receiver domain.

Growth kinetics data by McDaniels and Binder in 2012 suggested the possibility of having another metal binding domain elsewhere in the receptor apart from the ethylene binding domain.

This study ruled out the possibility of having the metal binding site in the receiver domain. In order to locate a metal binding site for any given protein, it is essential to have a clear understanding of its three dimensional structure. Unfortunately, only the receiver domain of ETR1 and CA domain of its kinase domain were crystallized and the information on the crystal structures of other domains in other receptors is very limited. This study facilitated visualization of other cytosolic domains of ethylene receptors.

The homology models created for kinase domains of ethylene receptors suggested that although the sub families function through different kinases, structurally they were similar to sensor histidine kinases. ERS2 had been an exception for this and the DHp domain of ERS2 is yet to be characterized well. This was partially due to the lack of proper templates that had close sequence similarity with ERS2 kinase domain. The structure of kinase domain in these receptors consisted of DHp domain with two large alpha helices towards the N-terminus which are highly conserved in all the receptors but with a slight variation due to the presence of a gap when aligned to the crystal structure. It was interesting to note that this gap persisted in all the receptors suggesting that this is a highly variable region and the model obtained a loop conformation along the length of the first alpha helix in the DHp domain. This is followed by CA subunit that has extremely conserved initial art of the protein consisting of one beta strand followed by an alpha helix and then by two beta strands and another alpha helix. The region after this is highly variable with different loop conformations.

There are not many GAF domains that have already been crystallized in PDB and hence the templates selected for GAF domains using PDB BLAST were very diverse. The models suggested that GAF domains mostly have conserved alpha helices and the models generated either long loops or very short beta strands against beta strands in the templates. Although the

function of GAF domains is yet to be determined, we observed that it could be a potential site for the second metal ion binding. We could arrive at this opinion because most of the templates suggested for GAF domain have metal binding sites and the sequence is rich in residues like E, D, H, Y , T, P which are known to bind positive metal ions, especially silver ions.

# 7 CONCLUSIONS AND FUTURE PROSPECTS

Ethylene receptors are highly redundant but have non-overlapping functions. Silver ions act as non-competitive inhibitors to ethylene receptors, support ethylene binding but block ethylene responses. Data presented in the earlier part of this thesis suggests that ETR1 receiver domain has little or no role in mediating ethylene responses to silver. Understanding how silver blocks ethylene responses could be of great help in horticultural industries to delay fruit ripening and increase shelf life of fruits and flowers. A remaining question would be the role of each of other receptor isoforms in mediating responses to silver.

Homology modeling of receiver domains suggested the importance of γ-loop in sub functionalization of ethylene receptors. The three dimensional structure would give future insights into understanding the receptor function better. Identification of possible metal ion binding site could be a major future prospect of this project. The second metal ion binding domain could be possibly in the GAF domain but this could be substantially proved once more GAF domains become available in PDB. The homology modeling project could be expanded endlessly to find different domains and mutate them and understand their physiological importance.

# REFERENCES

Abeles FB, Morgan PW, Saltveit ME Jr (1992) Ethylene in plant biology, 2nd edn. Academic Press, San Diego

Aloy P, Pichaud M, Russell RB. Protein complexes: structure prediction challenges for the 21st century. Current opinion in structural biology. 2005;15(1):15-22. Epub 2005/02/19.

Aravind L, Ponting CP. The GAF domain: an evolutionary link between diverse phototransducing proteins. Trends Biochem Sci. 1997;22:458–459

Bakshi A, Wilson RL, Lacey RF, Kim H, Wuppalapati SK, Binder BM. Identification of Regions in the Receiver Domain of the ETHYLENE RESPONSE1 Ethylene Receptor of Arabidopsis Important for Functional Divergence. Plant. Physiol. (in press)

Berg, A. R. and K. Peacock (1992). "Growth-Patterns in Nutating and Non nutating Sunflower (Helianthus-Annuus) Hypocotyls." American Journal of Botany79(1): 77-85.

Beyer, E.M. (1976). A potent inhibitor of ethylene action in plants. Plant Physiol. 58 268–271.

Binder BM, Chang C, Schaller GE. Perception of Ethylene by Plants – Ethylene Receptors. Annual Plant Reviews Volume 44: Wiley-Blackwell; 2012. p. 117-45.

Binder BM, O'malley RC, Wang WY, Moore JM, Parks BM, Spalding EP, Bleecker AB (2004b) Arabidopsis seedling growth response and recovery to ethylene: a kinetic analysis. Plant Physiol 136: 2913–2920

Binder BM, Rodriguez FI, Bleecker AB. The copper transporter RAN1 is essential for biogenesis of ethylene receptors in Arabidopsis. The Journal of biological chemistry. 2010;285(48):37263-70. Epub 2010/09/30.

Binder BM. Ethylene-Stimulated Nutations Do Not Require ETR1 Receptor Histidine Kinase Activity. Plant signaling & behavior. 2006;1(6):287-9. Epub 2006/11/01.

Binder, B. M., R. C. O'Malley, W. Wang, T. C. Zutz and A. B. Bleecker (2006). "Ethylene stimulates nutations that are dependent on the ETR1 receptor." Plant Physiol142(4): 1690-1700.

Blundell TL, Sibanda BL, Sternberg MJ, Thornton JM. Knowledge-based prediction of protein structures and the design of novel molecules. Nature. 1987;326(6111):347-52. Epub 1987/03/01.

Bourret RB. Receiver domain structure and function in response regulator proteins. Current opinion in microbiology. 2010;13(2):142-9. Epub 2010/03/10.

Burg SP, Burg EA. Molecular requirements for the biological activity of ethylene. Plant physiology. 1967;42(1):144-52. Epub 1967/01/01.

Chang, C., Kwok, S.F., Bleeker, A.B. &Meyerowitz, E.M. (1993).Arabidopsis ethylene-response gene ETR1: similarity of product totwo-component regulators. Science 262, 539-544.

Chen YF, Randlett MD, Findell JL, Schaller GE. Localization of the ethylene receptor ETR1 to the endoplasmic reticulum of Arabidopsis. Journal of Biological Chemistry. 2002;277(22):19861-6.

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31: 3497–3500

Greer J. Comparative model-building of the mammalian serine proteases. Journal of molecular biology. 1981;153(4):1027-42. Epub 1981/12/25.

Grefen, C., and Harter, K. (2004). Plant two-component systems: principles,functions, complexity and cross talk. Planta 219, 733–742.

Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences of the United States of America. 1992;89(22):10915-9. Epub 1992/11/15.

Hirayama T, Alonso JM. Ethylene captures a metal! Metal ions are involved in ethylene perception and signal transduction. Plant & cell physiology. 2000;41(5):548-55. Epub 2000/08/10.

Hirayama T, Kieber JJ, Hirayama N, Kogan M, Guzman P, Nourizadeh S, et al. RESPONSIVE-TO-ANTAGONIST1, a Menkes/Wilson disease-related copper transporter, is required for ethylene signaling in Arabidopsis. Cell. 1999;97(3):383-93. Epub 1999/05/13.

http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=FAQ#expect

Jeffery S. Thompson RLH, and John F. Whitney Copper( 1)-Olefin Complexes. Support for the Proposed Role of Copper in the Ethylene Effect in Plants. J Am Chem Soc 1983;105:3522-7.

Kim H, Helmbrecht EE, Stalans MB, Schmitt C, Patel N, Wen C-K, Wang W, Binder BM (2011) Ethylene receptor ETR1 domain requirements for ethylene responses in Arabidopsis seedlings. Plant Physiol 156: 417-429.

Labute, P.; Protonate3D: Assignment of Ionization States and Hydrogen Coordinates to Macromolecular Structures; Proteins 75 (2008) 187–205.

Mayerhofer H, Panneerselvam S, Kaljunen H, Tuukkanen A, Mertens HD, Mueller-Dieckmann J. Structural model of the cytosolic domain of the plant ethylene receptor 1 (ETR1). The Journal of biological chemistry. 2015;290(5):2644-58. Epub 2014/12/03.

Moussatche P, Klee HJ. Autophosphorylation activity of the Arabidopsis ethylene receptor multigene family. The Journal of biological chemistry. 2004;279(47):48734-41. Epub 2004/09/11.

Murashige, T. and Skoog, F. (1962) A revised medium for rapid growth and bioassays with tobacco tissue culture. Physiol. Plant. 15, 473-497.

O'Malley, R., and Bleecker, A.B. (2003). Ethylene perception in Arabidopsis by the ETR1 receptor family: evaluating a possible role for two-component signalling in plant ethylene responses. In Histidine Kinases in Signal Transduction, Inouye, M., and Dutta, R., eds (San Diego: Academic Press).

Parkinson JS, Kofoid EC. Communication modules in bacterial signaling proteins. Annual review of genetics. 1992;26:71-112. Epub 1992/01/01.

Parks, B.M., M.H. Cho, and E.P. Spalding, Two Genetically Separable Phases of Growth Inhibition Induced by Blue Light in Arabidopsis Seedlings. . Plant Physiology, 1998. 118: p. 609-615.

Petrey D, Chen TS, Deng L, Garzon JI, Hwang H, Lasso G, et al. Template-based prediction of protein function. Current opinion in structural biology. 2015;32C:33-8. Epub 2015/02/14.

Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic acids research. 2012;40(Database issue):D71-5. Epub 2011/11/22.

Rodriguez FI, Esch JJ, Hall AE, Binder BM, Schaller GE, Bleecker AB. A copper cofactor for the ethylene receptor ETR1 from Arabidopsis. Science. 1999;283(5404):996-8. Epub 1999/02/12.

Rodriguez, F. I., J. J. Esch, A. E. Hall, B. M. Binder, G. E. Schaller and A. B. Bleecker (1999). "A copper cofactor for the ethylene receptor ETR1 from Arabidopsis." Science 283(5404): 996-998.

Santner, A., L. I. Calderon-Villalobos and M. Estelle (2009). "Plant hormones are versatile chemical regulators of plant growth." Nat ChemBiol5(5): 301-307.

Schaller GE, Bleecker AB. Ethylene-binding sites generated in yeast expressing the Arabidopsis ETR1 gene. Science. 1995;270(5243):1809-11. Epub 1995/12/15.

Schaller GE, Ladd AN, Lanahan MB, Spanbauer JM, Bleecker AB. The ethylene response mediator ETR1 from Arabidopsis forms a disulfide-linked dimer. The Journal of biological chemistry. 1995;270(21):12526-30. Epub 1995/05/26.

Shakeel SN, Wang X, Binder BM, Schaller GE (2013) Mechanisms of signal transduction by ethylene: overlapping and nonoverlapping signalling roles in a receptor family. AoB Plants doi: 10.1093/aobpla/plt010: 1-16.

Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins. 1993;17(4):355-62. Epub 1993/12/01.

Stewart RC. Protein histidine kinases: assembly of active sites and their regulation in signaling pathways. Current opinion in microbiology. 2010;13(2):133-41. Epub 2010/02/02.

Sutcliffe MJ, Haneef I, Carney D, Blundell TL. Knowledge based modelling of homologous proteins, Part I: Three-dimensional frameworks derived from the simultaneous superposition of multiple structures. Protein engineering. 1987;1(5):377-84. Epub 1987/10/01.

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research. 1994;22(22):4673-4680.

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research. 1994;22(22):4673-4680.

Topham CM, McLeod A, Eisenmenger F, Overington JP, Johnson MS, Blundell TL. Fragment ranking in modelling of protein structure. Conformationally constrained environmental amino acid substitution tables. Journal of molecular biology. 1993;229(1):194-220. Epub 1993/01/05.

Tramontano A, Morea V. Assessment of homology-based predictions in CASP5. Proteins. 2003;53 Suppl 6:352-68. Epub 2003/10/28.

Ulijasz AT, Cornilescu G, von Stetten D, Cornilescu C, Velazquez Escobar F, Zhang J, et al. Cyanochromes are blue/green light photoreversible photoreceptors defined by a stable double cysteine linkage to a phycoviolobilin-type chromophore. The Journal of biological chemistry. 2009;284(43):29757-72. Epub 2009/08/13.

Voet-van-Vormizeele J, Groth G. Ethylene controls autophosphorylation of the histidine kinase domain in ethylene receptor ETR1. Molecular plant. 2008;1(2):380-7. Epub 2008/03/01.

Wilson RL , Lacey RF, Binder BM. Ethylene Receptors - Biochemical Events. In: Wen C-K, editor. Ethylene in Plants: Springer; 2015. p. 45-60.

Wilson RL, Bakshi A and Binder BM (2014) Loss of the ETR1 ethylene receptor reduces the inhibitory effect of far-red light and darkness on seed germination of Arabidopsis thaliana. Front. Plant Sci. 5:433. doi: 10.3389/fpls.2014.00433

Wilson RL, Kim H, Bakshi A, Binder BM (2014b) The ethylene receptors ETHYLENE RESPONSE1 and ETHYLENE RESPONSE2 have contrasting roles in seed germination of Arabidopsis during salt stress. Plant Physiol 165: 1353-1366

Wolanin PM, Thomason PA, Stock JB. Histidine protein kinases: key signal transducers outside the animal kingdom. Genome biology. 2002;3(10):REVIEWS3013. Epub 2002/10/10.

Xie, F., Q. Liu, and C.-K. Wen, Receptor signal output mediated by the ETR1 N-terminus is primarily subfamily I receptor dependent. Plant Physiology, 2006. 142: p. 492-508.

Zhang L, Skolnick J. What should the Z-score of native protein structures be? Protein science : a publication of the Protein Society. 1998;7(5):1201-7. Epub 1998/05/30.

# APPENDIX

## Figures

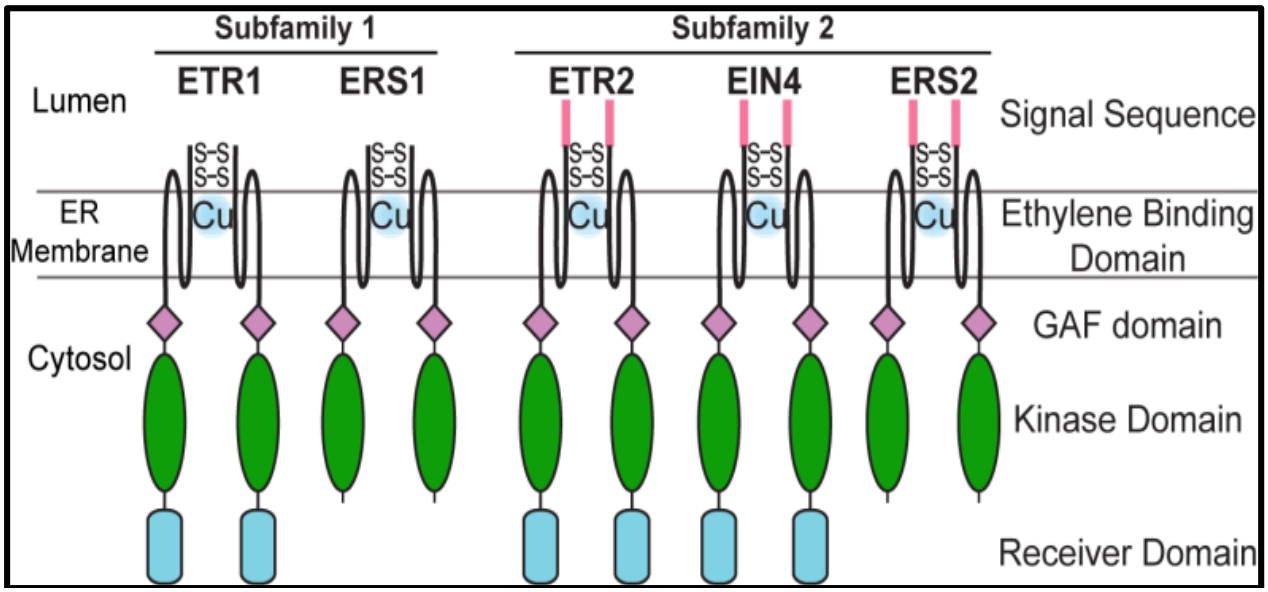



**Figure 1. Structure and classification of five isoforms of Ethylene receptor subfamilies in *Arabidopsis thaliana*.** (Lacey and Binder., 2014).

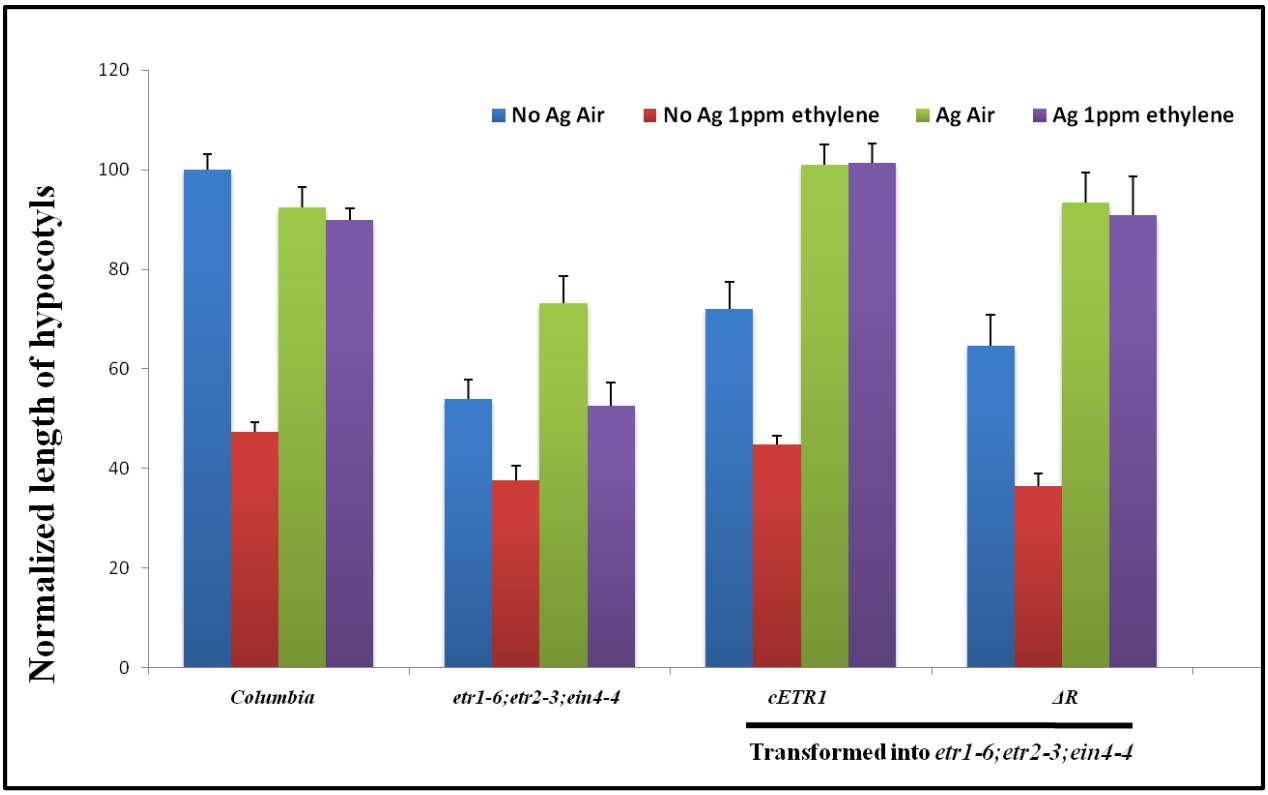



**Figure 2. Role of receiver domain in eliciting ethylene growth responses on etiolated *Arabidopsis* seedlings in the presence of silver ions.** Seedlings were grown in darkness for 4 days in air and 100 PPM ethylene on agar plates supplemented with 100µM silver nitrate. Hypocotyl growth of ethylene receptor loss-of-function mutant seedlings was compared to that of wildtype seedlings that were used as controls. *etr1-6;etr2-3;ein4-4* seedlings were transformed with cDNA transgene and the other triple loss-of-function mutant lacked the receiver domain at the end of C terminus. Supplementation and absence of ethylene and silver ions is clearly marked and represented.

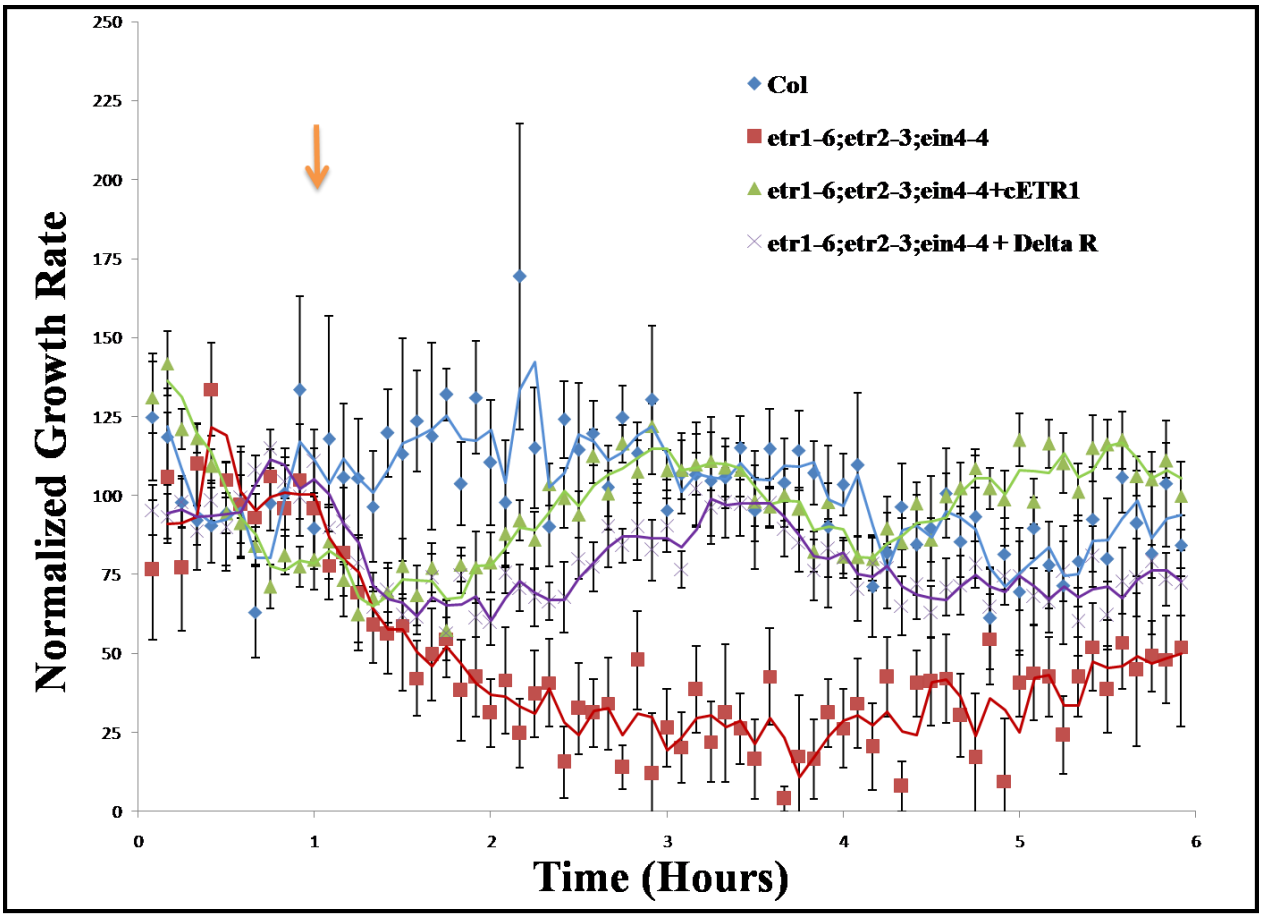



**Figure 3. Growth inhibition kinetics of ETR1 receiver domain in the presence of silver ions.** The seedlings were grown in air for one hour followed by addition of 1µL $L^{-1}$ ethylene for five hours. Onset of ethylene is indicated by arrow. The seedlings were grown on agar plates containing 100µM silver nitrate. The growth of seedlings was captured every 5 minutes in the time lapse set up and graphs were plotted. The graphs represent time on X-axis and normalized growth rate of hypocotyls on Y-axis. The data represents mean ± SE from at least 4 separate experiments.

```
CLUSTAL 2.1 multiple sequence alignment


ETR2_RD         QVLLVDTNDSNRAVTRKLLEKLGCDVTAVSSGFDCLTAIAPGSSSPSTSFQVVVLDLQMA 60
EIN4_RD         RITLADDDDVNRTVTKRLLEKLGCEVTAVSSGFECLNALS----NVEMSYRVVILDLQMP 56
ETR1_RD         KVLVMDENGVSRMVTKGLLVHLGCEVTTVSSNEECLRVVS-------HEHKVVFMDVCMP 53
                :: : * :. .* **: ** :***:**:***. :** .::        ..:**.:*: *.

ETR2_RD         EMDGYEVAMRIR------SRSWPLIVATTVSLDEEMWDKCAQIGINGVVRKPVVLRAMES 114
EIN4_RD         EMDGFEVAMKIRKFC---GHHWPLIIALTASTEDHVRERCLQMGMNGMIQKPVLLHVMAS 113
ETR1_RD         GVENYQIALRIHEKFTKQRHQRPLLVALSGNTDKSTKEKCMSFGLDGVLLKPVSLDNIRD 113
                 ::.:::*::*:       :  **::* : . :.   ::* .:*::*:: *** *  : .

ETR2_RD         ELRRVLLQADQLL-- 127
EIN4_RD         ELRRALQTASE---- 124
ETR1_RD         VLSDLLEPRVLYEGM 128
```

| **Name of the Receptor** | **% Identity** | | |
|---|---|---|---|
| | ETR1_RD | ETR2_RD | EIN4_RD |
| ETR1_Receiver Domain | 100.00 | 55.37 | 35.83 |
| ETR2_Receiver Domain | | 100.00 | 36.36 |
| EIN4_Receiver Domain | | | 100.00 |

**Figure 4. Sequence alignment and identity matrix of receiver domains of Ethylene receptors.**

**A.** Sequence alignment of the receiver domain of the three ethylene receptors in *Arabidopsis* using CLUSTAL W. The residues with an asterisk (*) denote conserved residues, colons (**:**) denote conserved substitutions and periods (**.**) denote semi conserved substitutions (Chenna *et al.*,2003). **B.** The identity matrix was created using the same program which shows the percentage of conserved residues among the receptors.

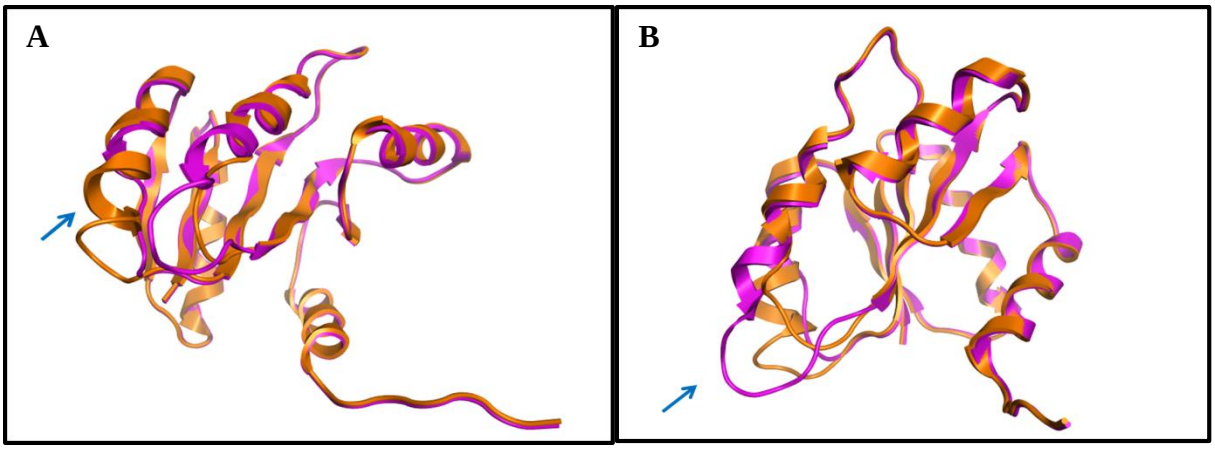



**Figure 5. Three dimensional structure of ETR2 receiver domain using 1DCF.A as template.**

A, B. Structurally diverged regions that were conformationally different in the model from respective regions in the template, were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

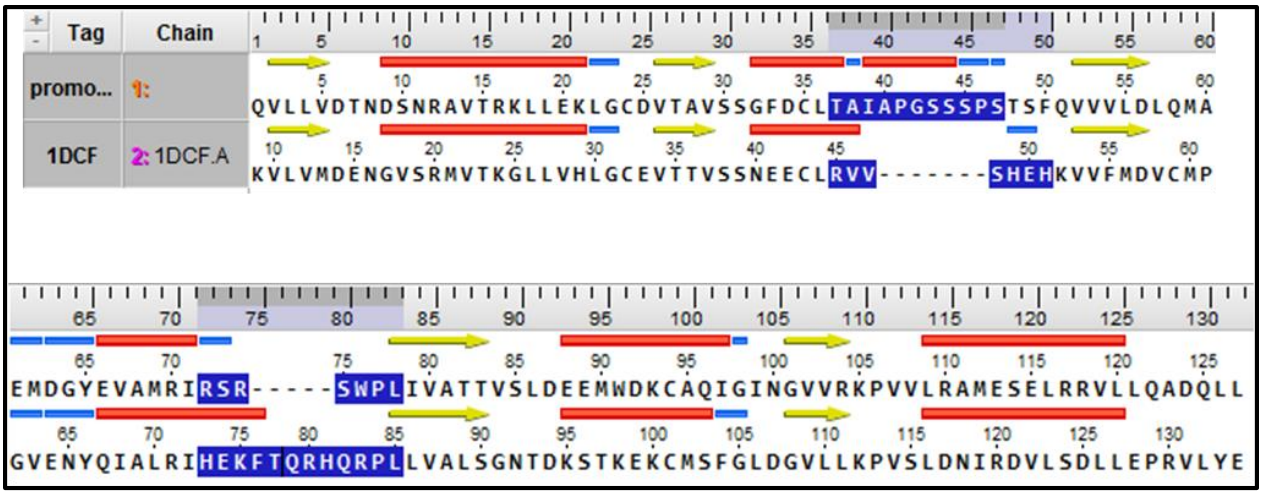



**Figure 6. Structurally diverged regions in target sequence, ETR2 receiver domain in comparison to template, 1DCF.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bar, beta strands were represented as yellow arrows, and bends were represented as short blue lines. Target sequence was denoted as Promodel 4 and template as 1DCF.A.

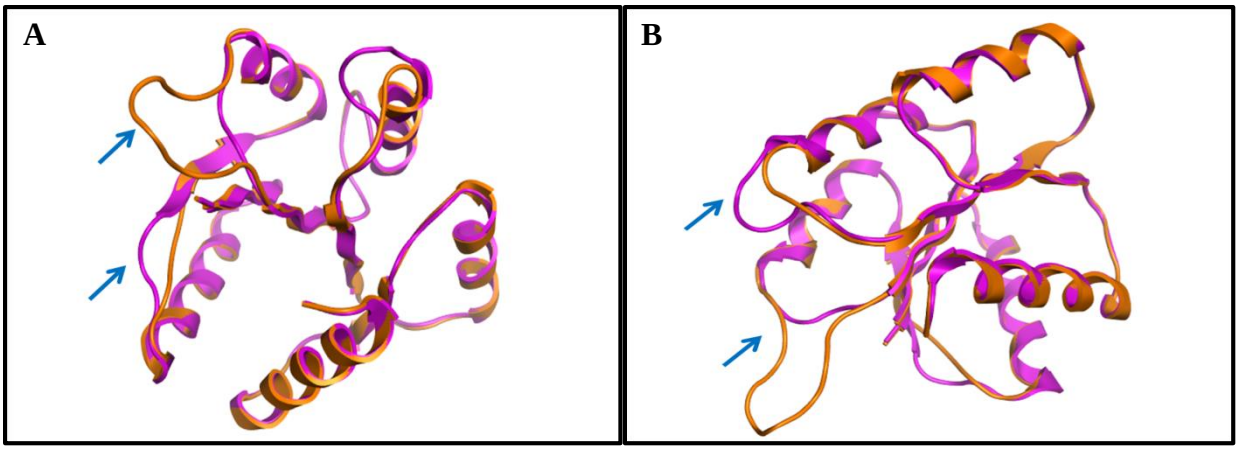



**Figure 7. Three dimensional structure of ETR2 receiver domain using 3RVK.A as template.**

**A,B.** Structurally diverged regions in the model that were from respective regions in the template, were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

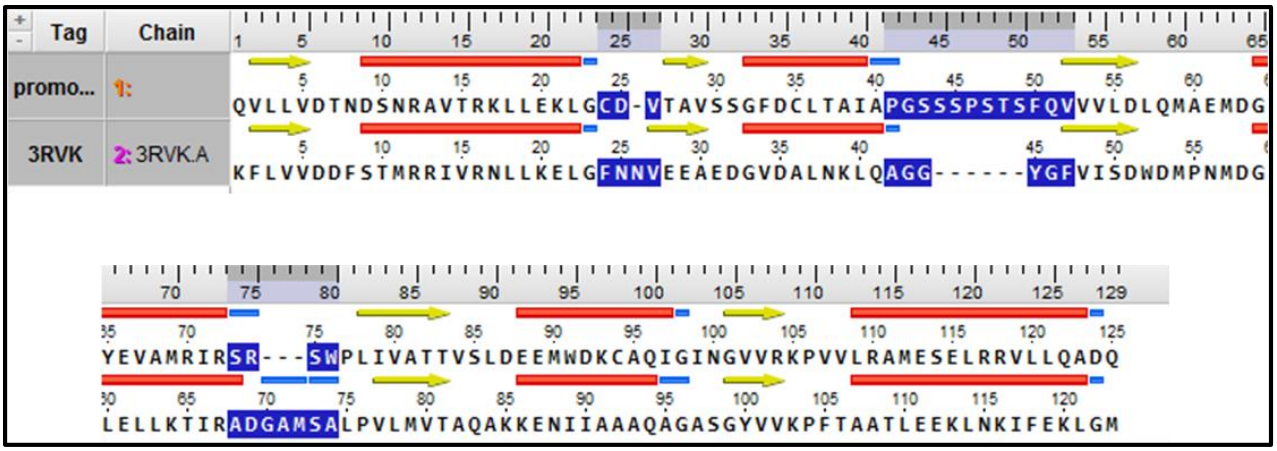



**Figure 8. Structurally diverged regions in target sequence, ETR2 receiver domain in comparison to template, 3RVK.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bar, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel 1 and template as 3RVK.A.

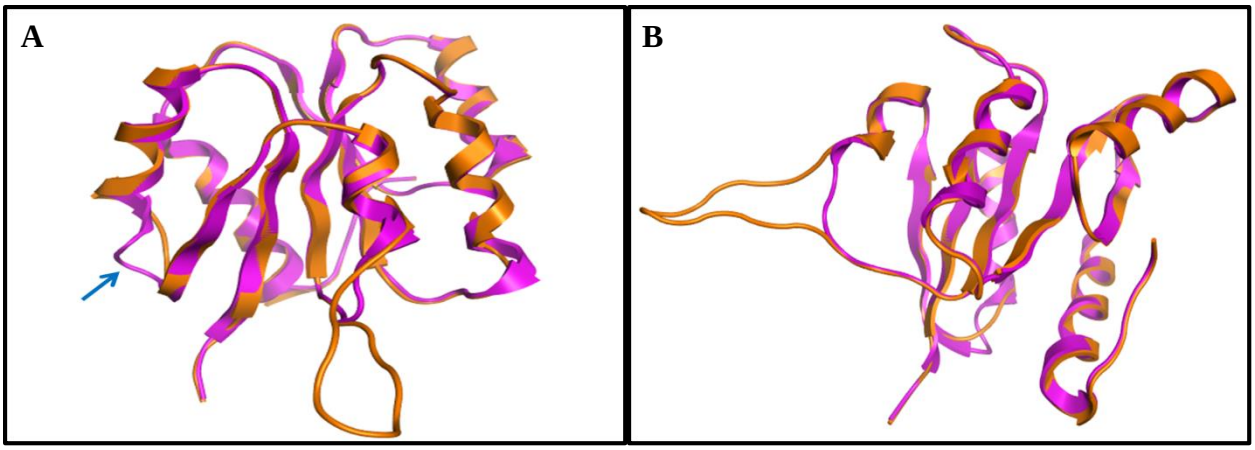



**Figure 9. Three dimensional structure of ETR2 receiver domain using 1P2F.A. as template.** A,B. Structurally diverged regions in the model that were from respective regions in the template, were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

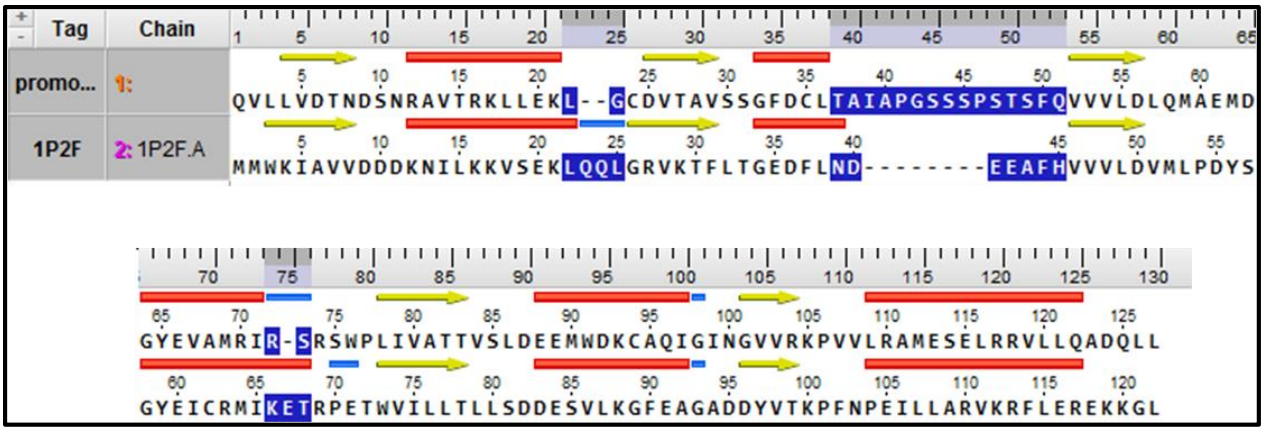



**Figure 10. Structurally diverged regions in target sequence, ETR2 receiver domain in comparison to template. 1P2F.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bar, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel 5 and template as 1P2F.A.

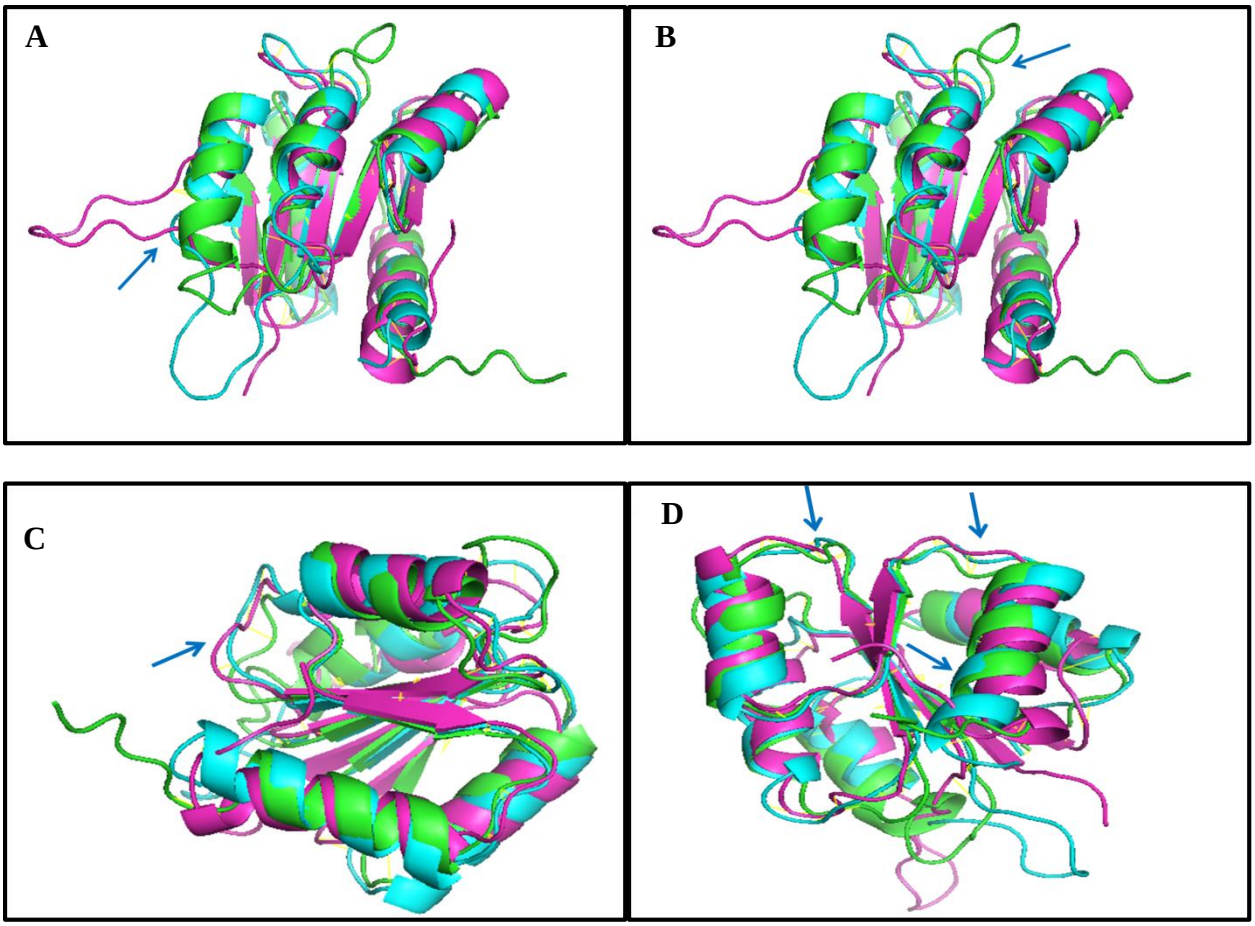



**Figure 11. Superimposing the possible models generated for ETR2 receiver domain.**
Models with lowest RMSD values generated against the templates 1DCF.A (represented in green), 3RVK.A (represented in blue) and 1P2F.A (represented in pink). were superimposed in PyMol and the structurally diverged regions were marked by blue arrows.

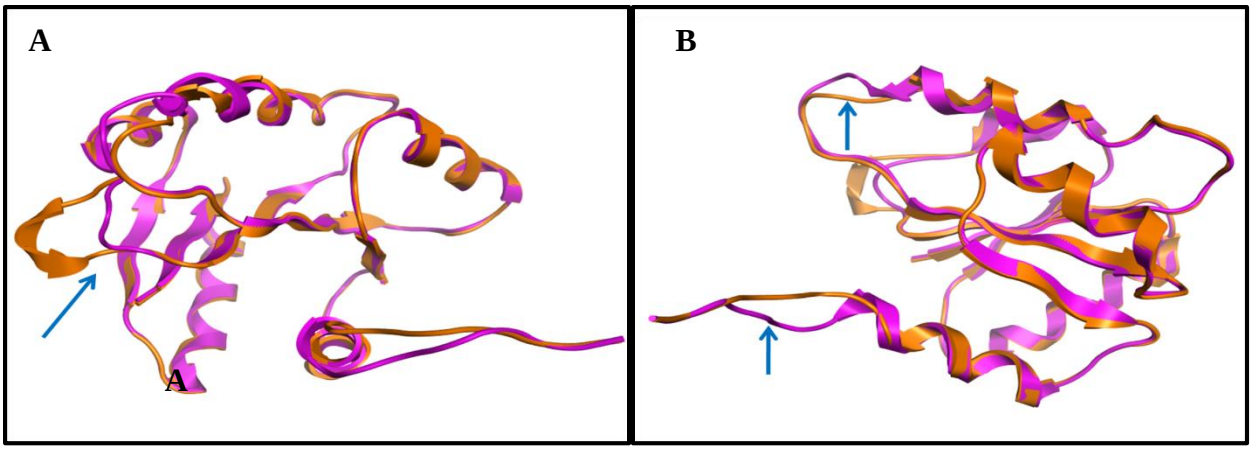



**Figure 12. Three dimensional structure of EIN4 receiver domain using 1DCF.A. as template.**

**A,B.** Structurally diverged regions in the model that were from respective regions in the template, were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

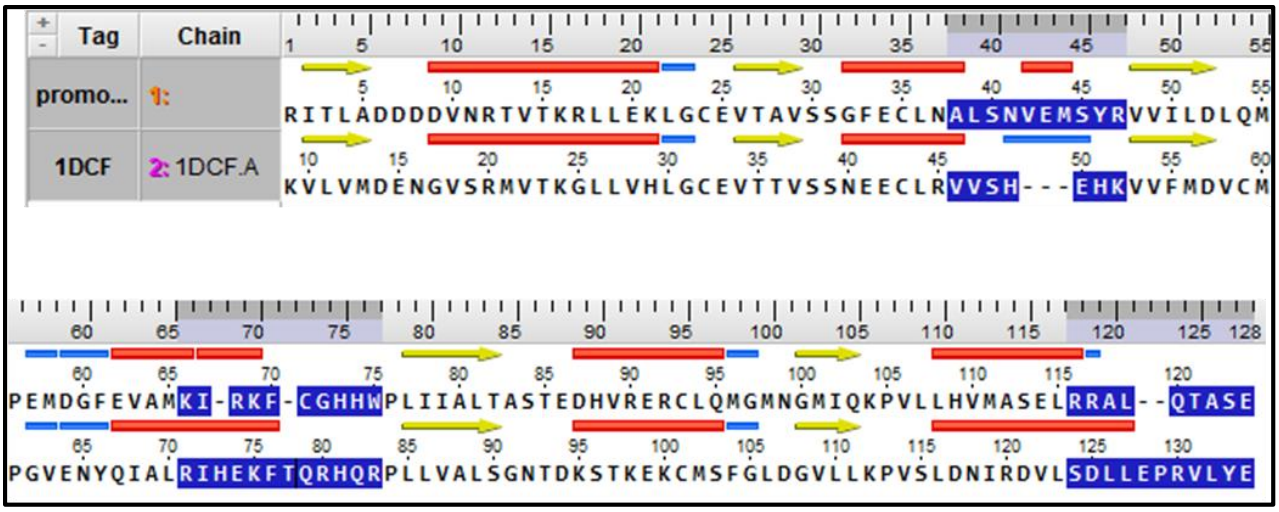



**Figure 13. Structurally diverged regions in target sequence EIN4 receiver domain in comparison to template, 1DCF.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bar, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel 1 and template as 1DCF.A.

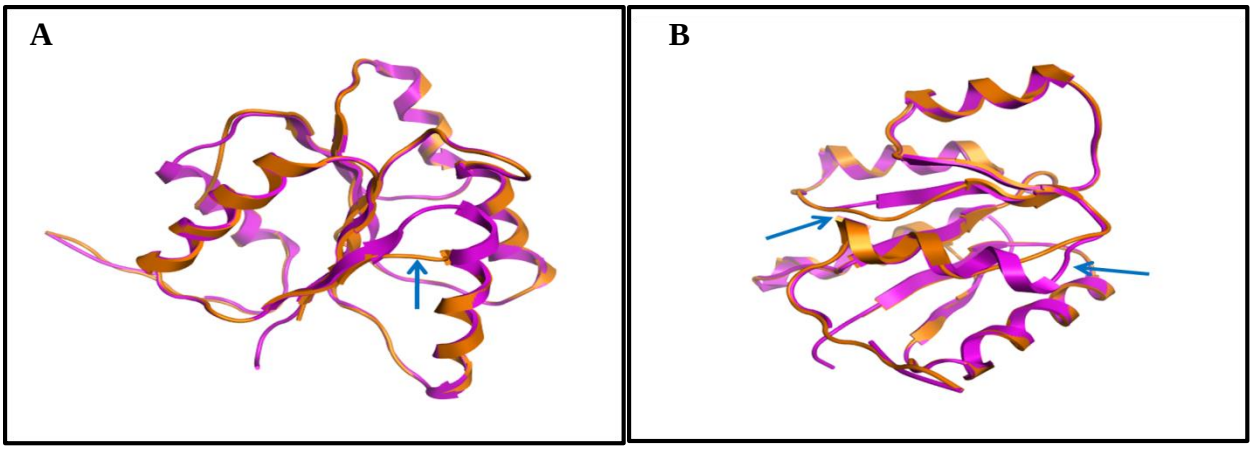



**Figure 14. Three dimensional structure of EIN4 receiver domain using 2R25.B. as template A,B.** Structurally diverged regions in the model that were from respective regions in the template, were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

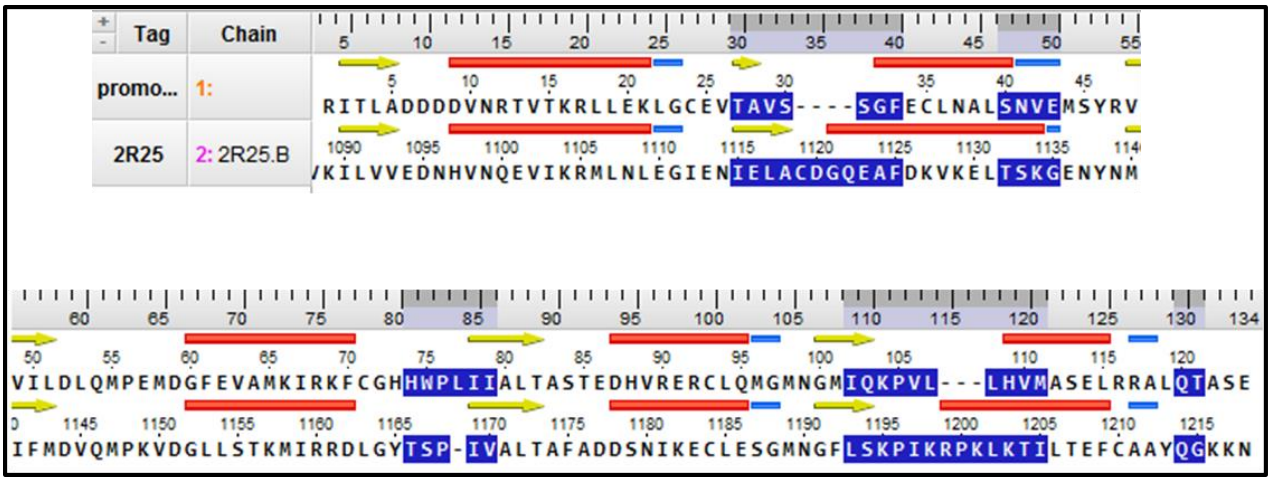



**Figure 15. Structurally diverged regions in target sequence EIN4 receiver domain in comparison to template, 2R25.B.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bar, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel 1 and template as 2R25.B.

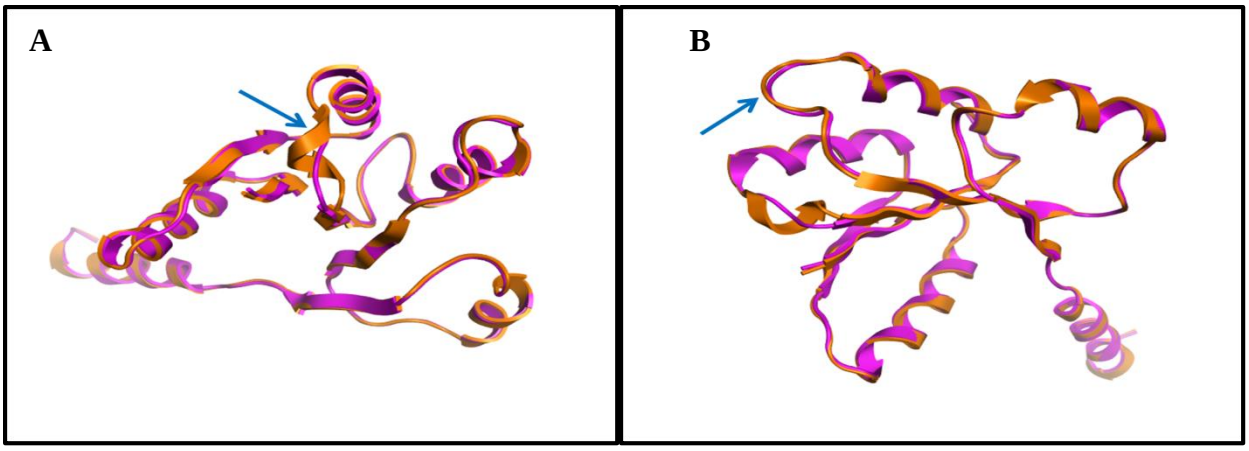



**Figure 16. Three dimensional structure of EIN4 receiver domain using 3C3M.A as template**

**A,B.** Structurally diverged regions in the model that were from respective regions in the template, were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

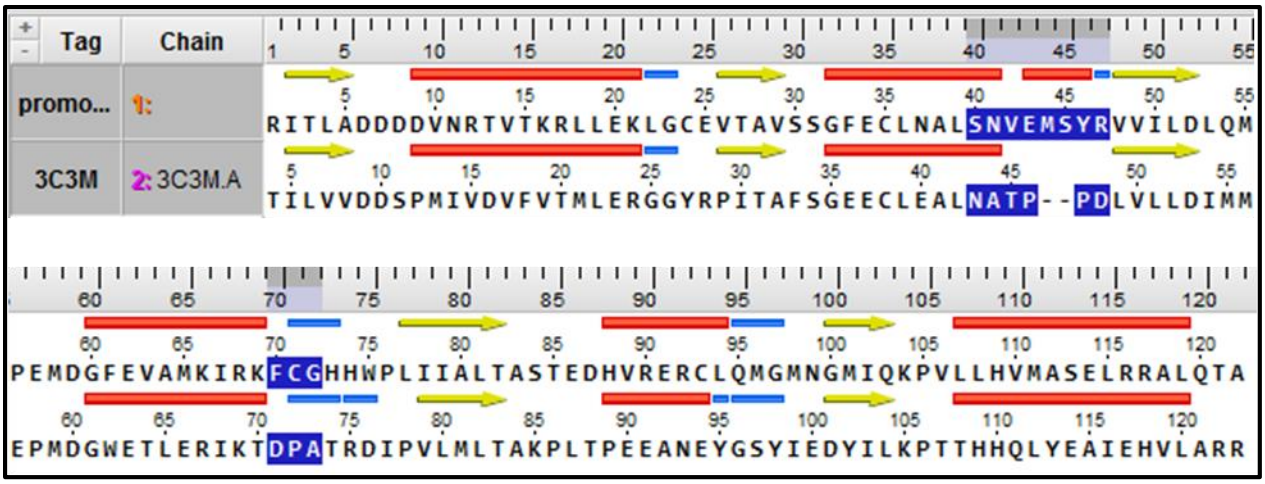



**Figure 17. Structurally diverged regions in target sequence EIN4 receiver domain in comparison to template 3C3M.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bar, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel 5 and template as 3C3M.A.

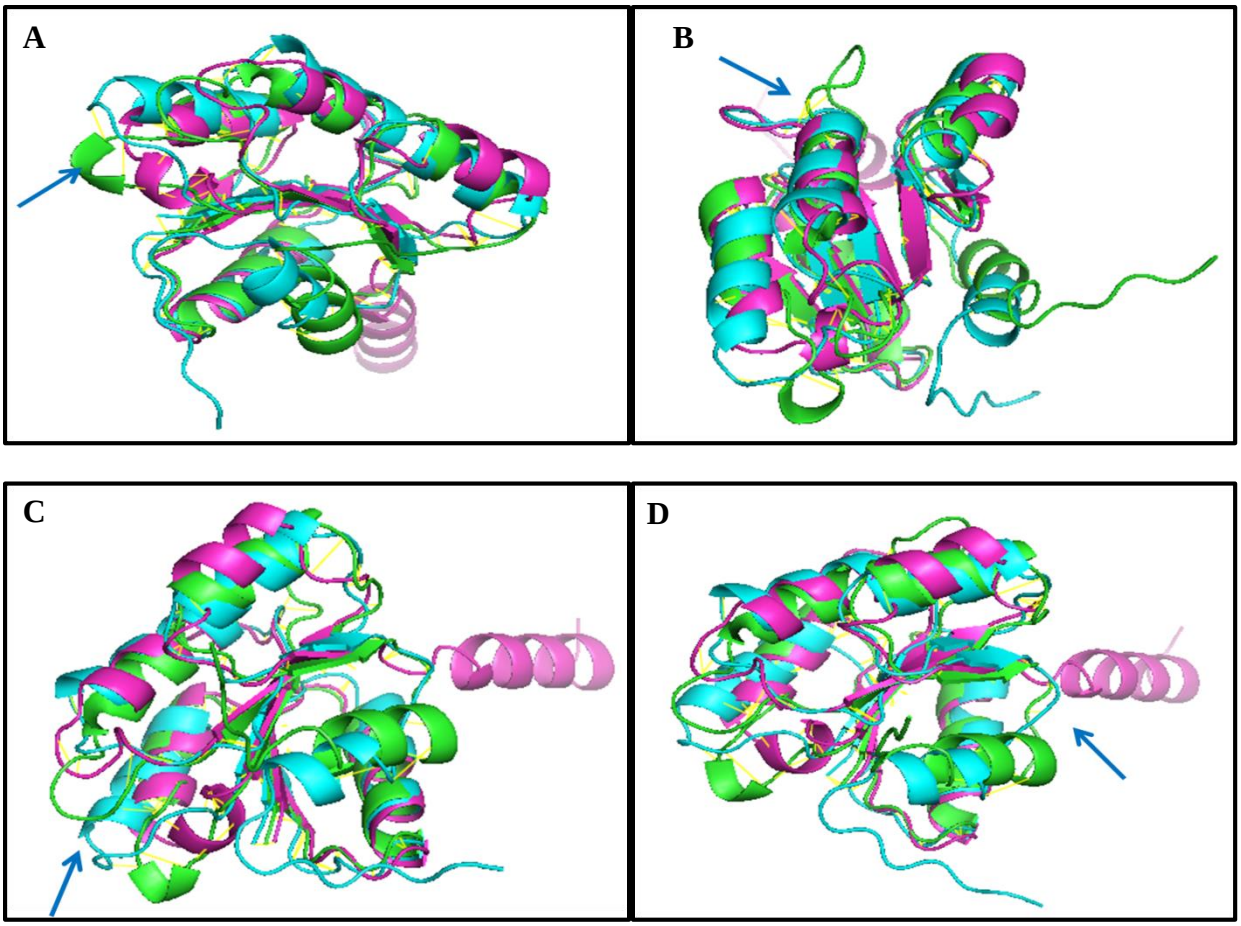



**Figure 18. Superimposing the possible models generated for EIN4 receiver domain.**
Models with lowest RMSD values generated against the templates 1DCF.A (represented in green), 2R25.B (represented in blue) and 3C3M.A (represented in pink). were superimposed in PyMol and the structurally diverged regions were marked by blue arrows.

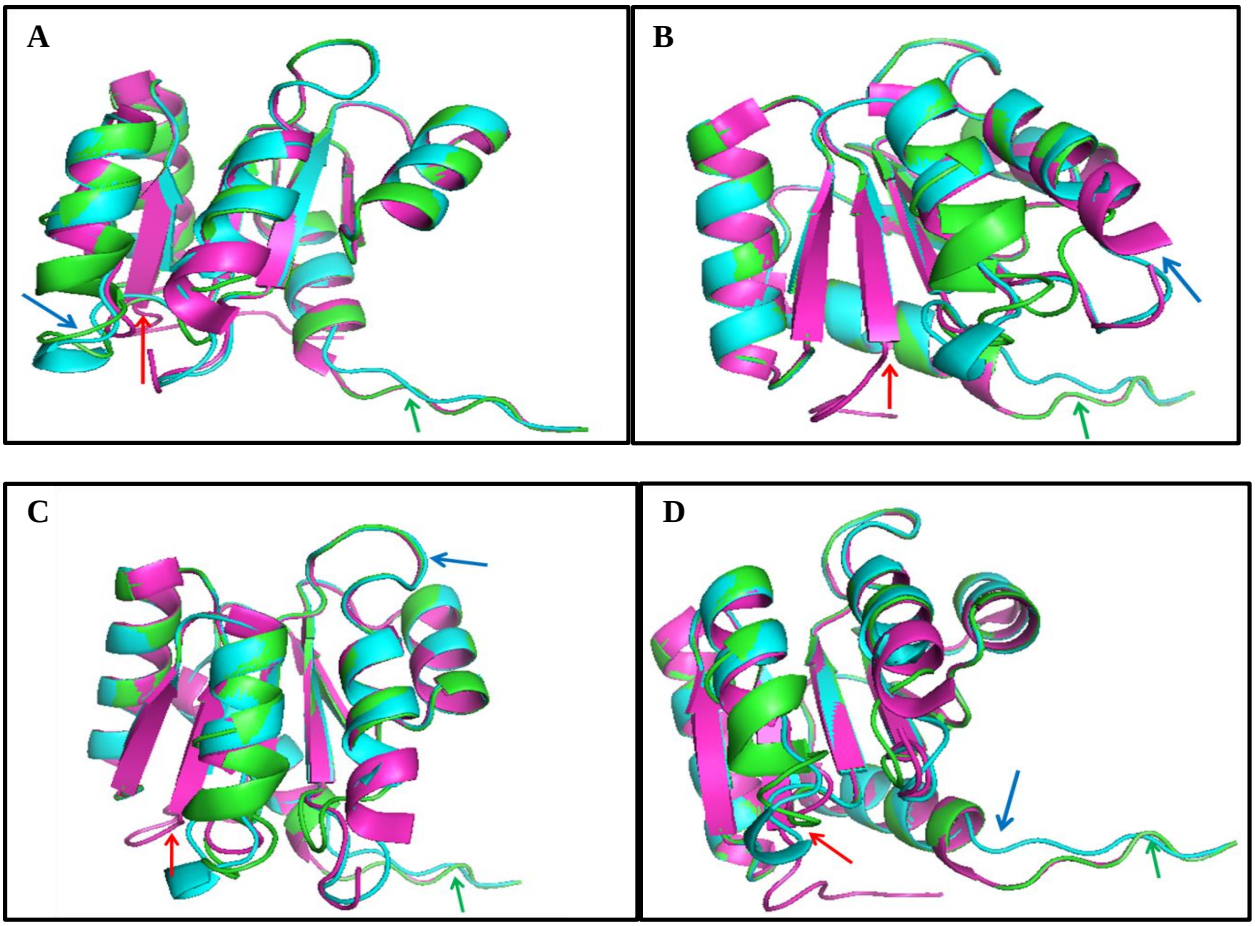
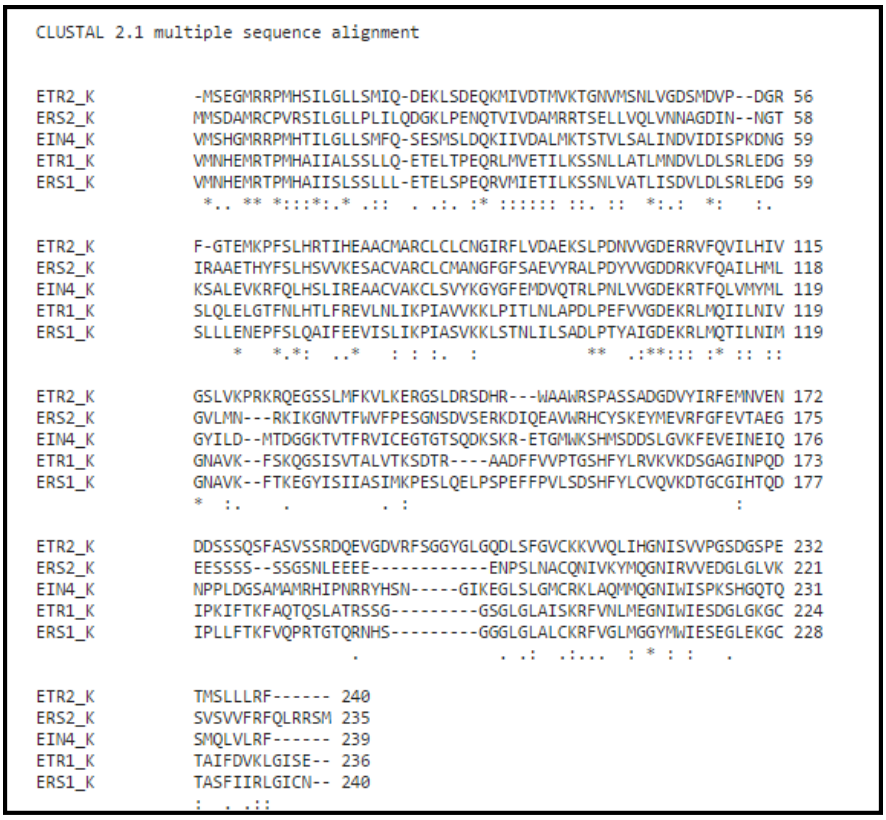



**Figure 19. Superimposing the possible models generated for ETR2 and EIN4 receiver domain against 1DCF.A.**

Models with lowest RMSD values generated against the templates 1DCF.A (represented in purple), for ETR2 receiver domain (represented in green) and EIN4 receiver domain (represented in blue) were superimposed in PyMol and the structurally diverged regions were marked by blue arrows.



```
CLUSTAL 2.1 multiple sequence alignment


ETR2_K          -MSEGMRRPMHSILGLLSMIQ-DEKLSDEQKMIVDTMVKTGNVMSNLVGDSMDVP--DGR 56
ERS2_K          MMSDAMRCPVRSILGLLPLILQDGKLPENQTVIVDAMRRTSELLVQLVNNAGDIN--NGT 58
EIN4_K          VMSHGMRRPMHTILGLLSMFQ-SESMSLDQKIIVDALMKTSTVLSALINDVIDISPKDNG 59
ETR1_K          VMNHEMRTPMHAIIALSSLLQ-ETELTPEQRLMVETILKSSNLLATLMNDVLDLSRLEDG 59
ERS1_K          VMNHEMRTPMHAIISLSSLLL-ETELSPEQRVMIETILKSSNLVATLISDVLDLSRLEDG 59
                 *.. ** *:::*:.* .::  . .:. :* :::::: ::. ::  *:.:  *:   :.

ETR2_K          F-GTEMKPFSLHRTIHEAACMARCLCLCNGIRFLVDAEKSLPDNVVGDERRVFQVILHIV 115
ERS2_K          IRAAETHYFSLHSVVKESACVARCLCMANGFGFSAEVYRALPDYVVGDDRKVFQAILHML 118
EIN4_K          KSALEVKRFQLHSLIREAACVAKCLSVYKGYGFEMDVQTRLPNLVVGDEKRTFQLVMYML 119
ETR1_K          SLQLELGTFNLHTLFREVLNLIKPIAVVKKLPITLNLAPDLPEFVVGDEKRLMQIILNIV 119
ERS1_K          SLLLENEPFSLQAIFEEVISLIKPIASVKKLSTNLILSADLPTYAIGDEKRLMQTILNIM 119
                    *   *.*:  ..*   : : :.  :           **  .:**::: :* :: ::

ETR2_K          GSLVKPRKRQEGSSLMFKVLKERGSLDRSDHR---WAAWRSPASSADGDVYIRFEMNVEN 172
ERS2_K          GVLMN---RKIKGNVTFWVFPESGNSDVSERKDIQEAVWRHCYSKEYMEVRFGFEVTAEG 175
EIN4_K          GYILD--MTDGGKTVTFRVICEGTGTSQDKSKR-ETGMWKSHMSDDSLGVKFEVEINEIQ 176
ETR1_K          GNAVK--FSKQGSISVTALVTKSDTR----AADFFVVPTGSHFYLRVKVKDSGAGINPQD 173
ERS1_K          GNAVK--FTKEGYISIIASIMKPESLQELPSPEFFPVLSDSHFYLCVQVKDTGCGIHTQD 177
                *  :.    .         . :                                 :

ETR2_K          DDSSSQSFASVSSRDQEVGDVRFSGGYGLGQDLSFGVCKKVVQLIHGNISVVPGSDGSPE 232
ERS2_K          EESSSS--SSGSNLEEEE------------ENPSLNACQNIVKYMQGNIRVVEDGLGLVK 221
EIN4_K          NPPLDGSAMAMRHIPNRRYHSN-----GIKEGLSLGMCRKLAQMMQGNIWISPKSHGQTQ 231
ETR1_K          IPKIFTKFAQTQSLATRSSG---------GSGLGLAISKRFVNLMEGNIWIESDGLGKGC 224
ERS1_K          IPLLFTKFVQPRTGTQRNHS---------GGGLGLALCKRFVGLMGGYMWIESEGLEKGC 228
                              .             . .:  .:...  : * : :   .

ETR2_K          TMSLLLRF------ 240
ERS2_K          SVSVVFRFQLRRSM 235
EIN4_K          SMQLVLRF------ 239
ETR1_K          TAIFDVKLGISE-- 236
ERS1_K          TASFIIRLGICN-- 240
                :  . .::
```

| Name of the Receptor | % Identity | | | | |
|---|---|---|---|---|---|
| | ETR1_K | ERS1_K | ETR2_K | ERS2_K | EIN4_K |
| ETR1_Kinase Domain | 100 | 64.41 | 25.78 | 20.18 | 27.26 |
| ERS1_ Kinase Domain | | 100 | 24.02 | 17.24 | 24.68 |
| ETR2_Kinase Domain | | | 100 | 40.36 | 37.77 |
| ERS2_Kinase Domain | | | | 100 | 33.04 |
| EIN4_Kinase Domain | | | | | 100 |

**Figure 20. Sequence alignment and identity matrix for kinase domains in Ethylene receptors.** Sequence alignment of the kinase domain of the three ethylene receptors in *Arabidopsis* using CLUSTAL W2. The residues with an asterisk (*) denote conserved residues, colons (:) denote conserved substitutions and periods (.) denote semi conserved substitutions (Chenna *et al.*, 2003). B. The identity matrix was created using the same program which shows the percentage of conserved residues among the receptors.

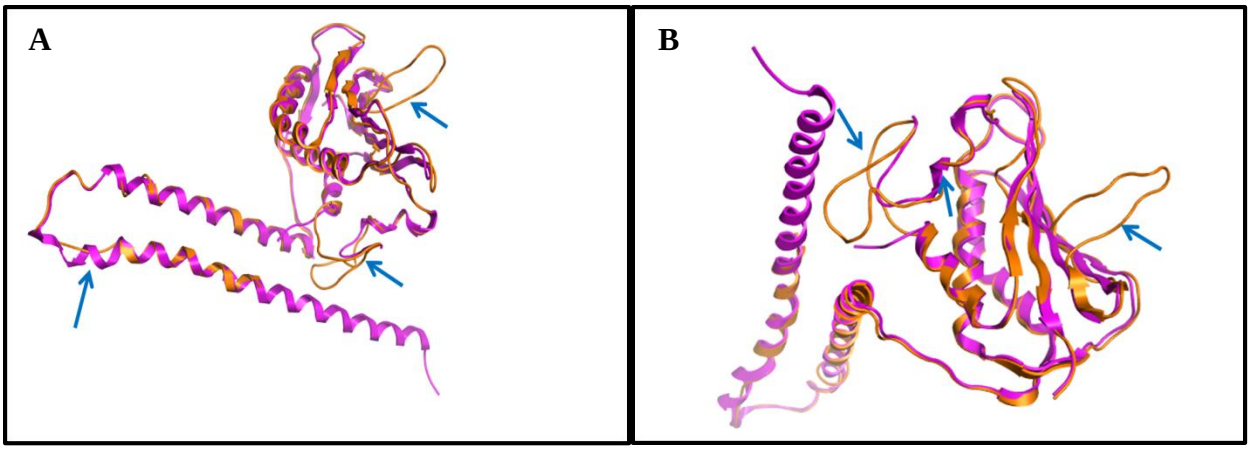



**Figure 21. Three dimensional structure of ETR1 kinase domain using 2C2A.A as template**

A,B. Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

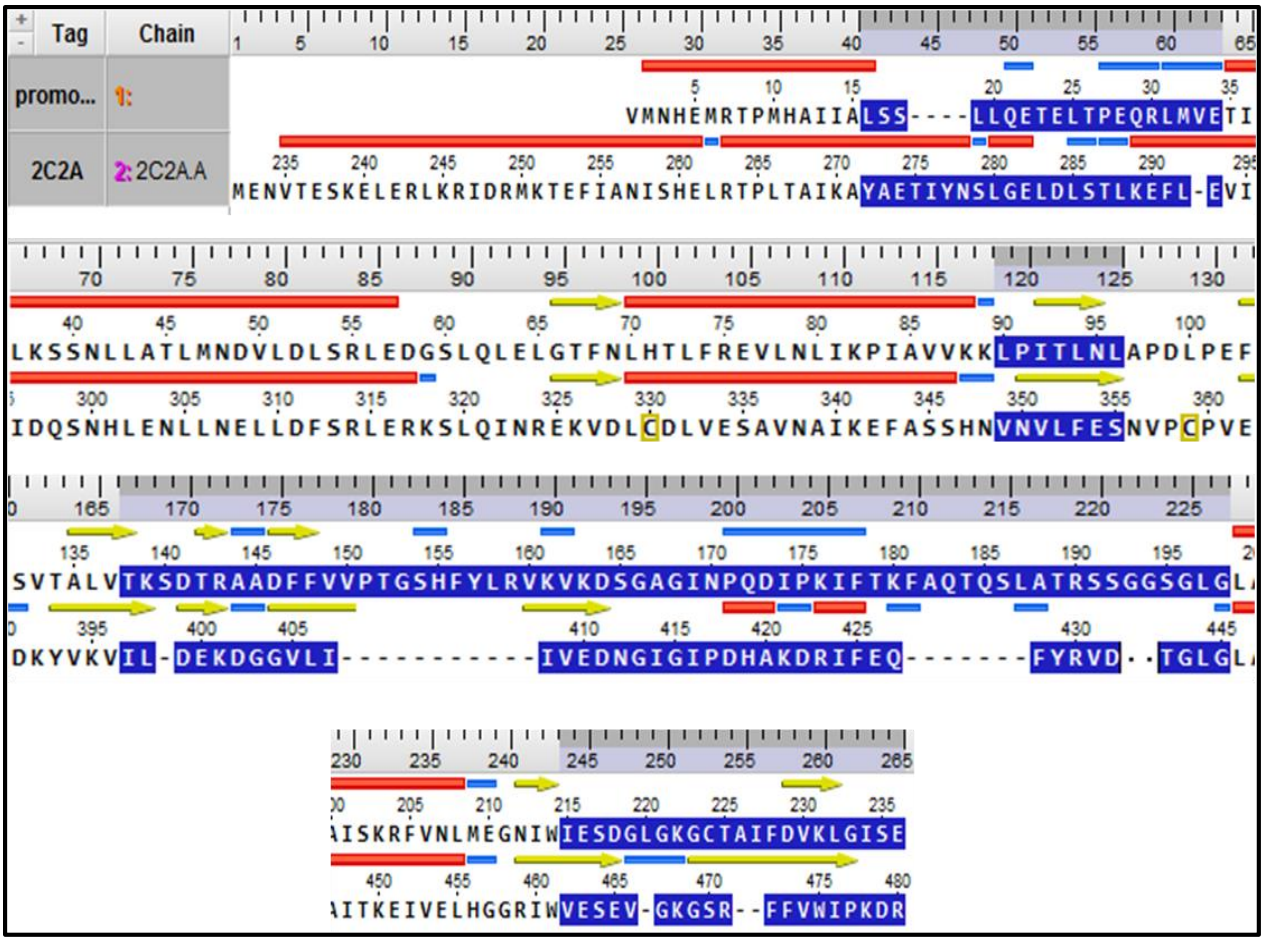



**Figure 22. Structurally diverged regions in target sequence, ETR1 kinase domain in comparison to template 2C2A.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel 9 and template as 2C2A.A.

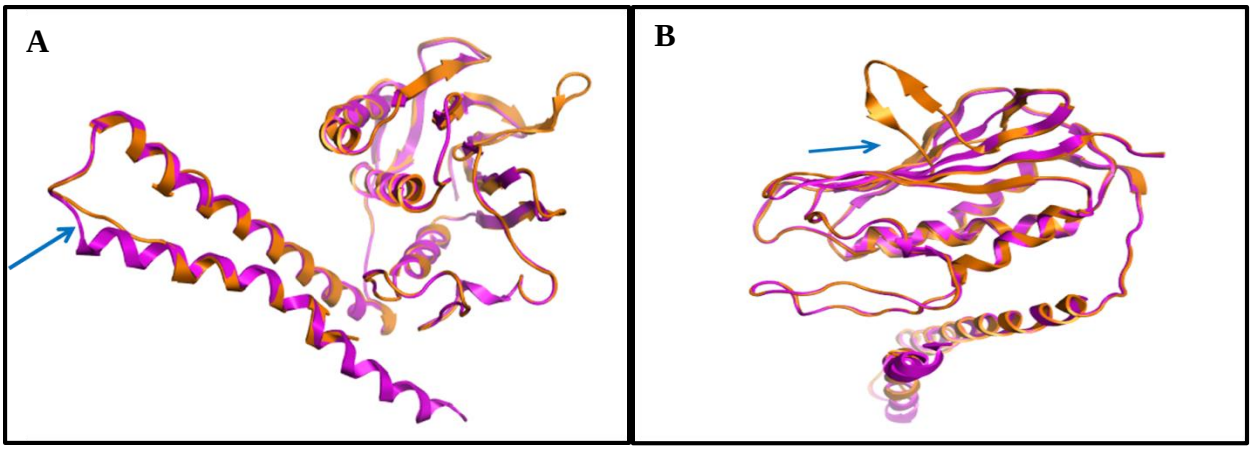



**Figure 23. Three dimensional structure of ETR1 kinase domain using 3DGE.A as template**

A,B. Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

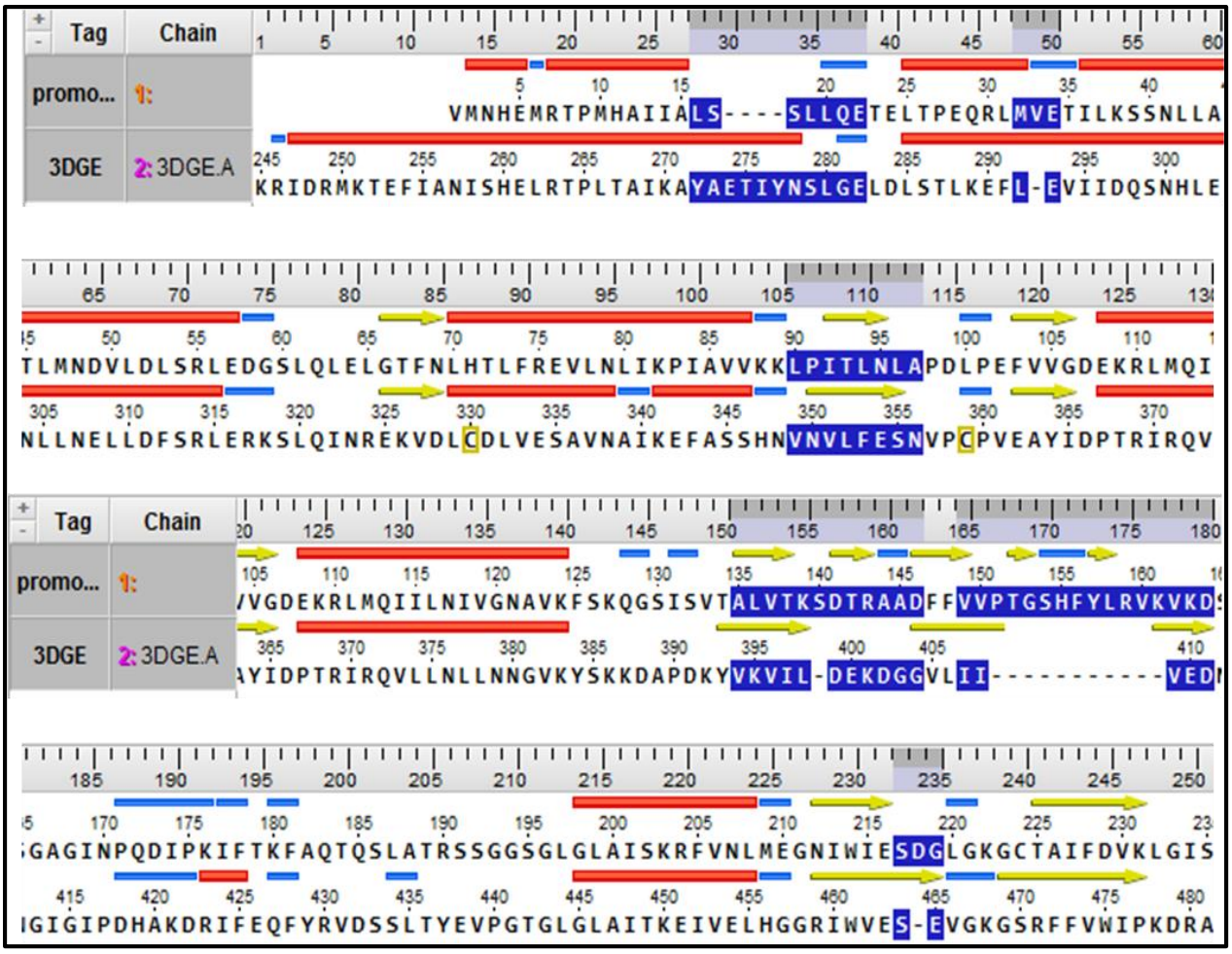



**Figure 24. Structurally diverged regions in target sequence, ETR1 kinase domain in comparison to template 3DGE.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel 4 and template as 3DGE.A.

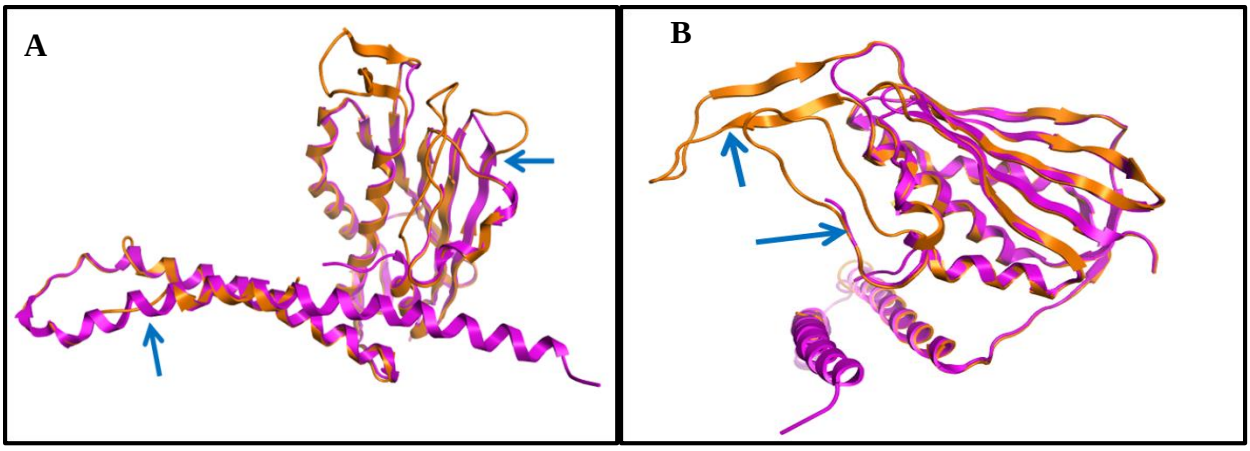



**Figure 25. Three dimensional structure of ERS1 kinase domain using 2C2A. A as template**

A,B. Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

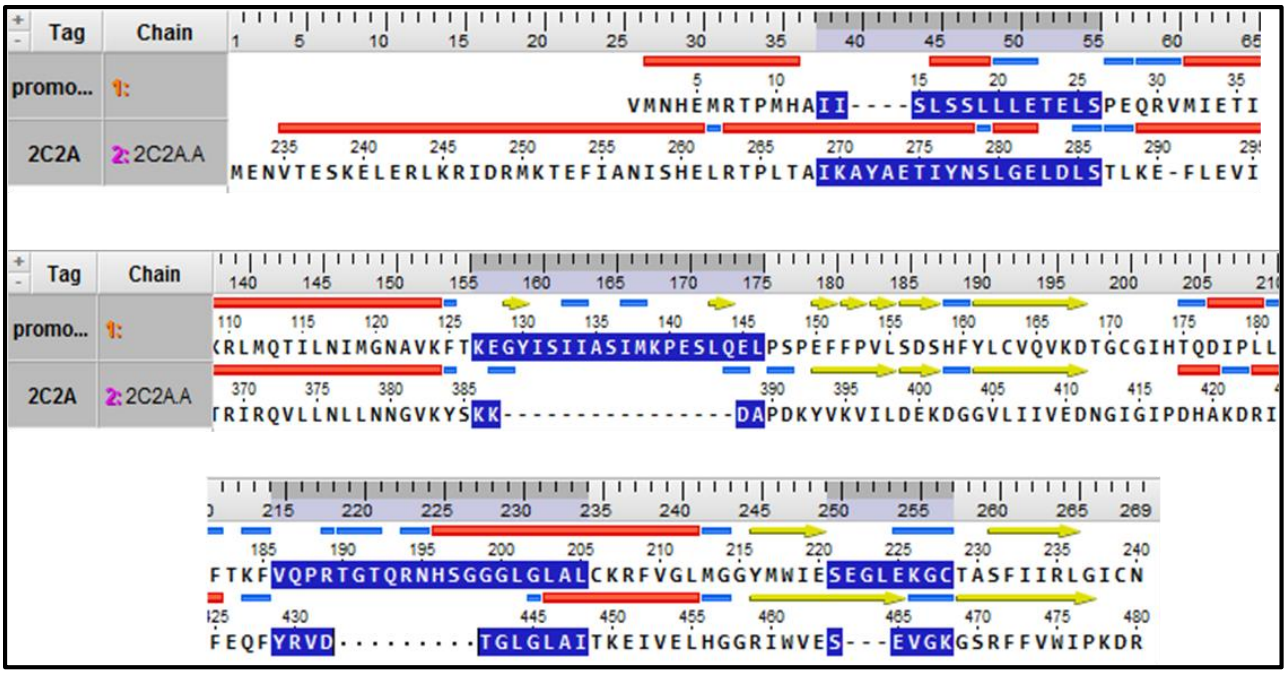



**Figure 26. Structurally diverged regions in target sequence ERS1 kinase domain in comparison to template 2C2A.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel 3 and template as 2C2A.A.

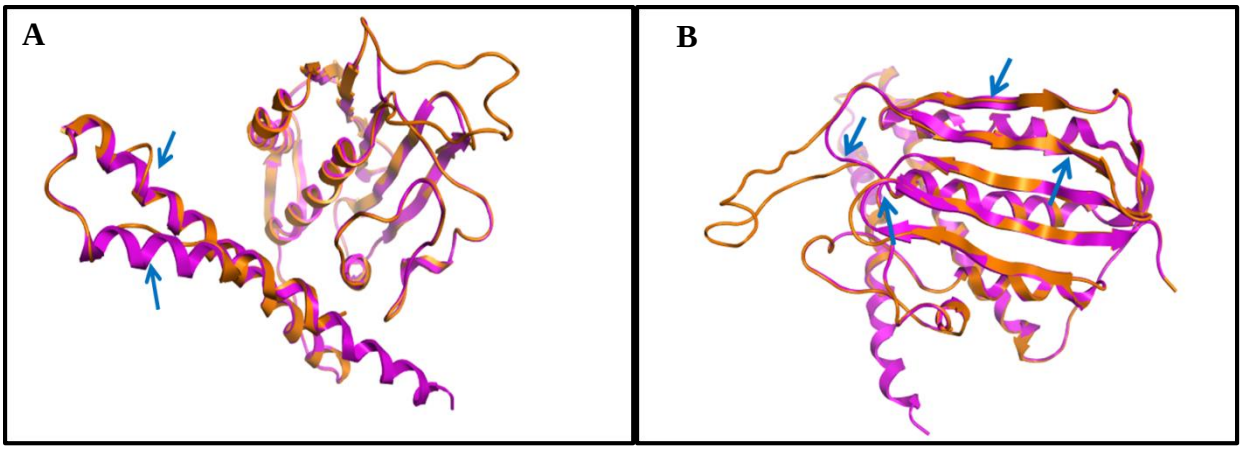



**Figure 27. Three dimensional structure of ERS1 kinase domain using 3DGE. A as template A,B.** Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

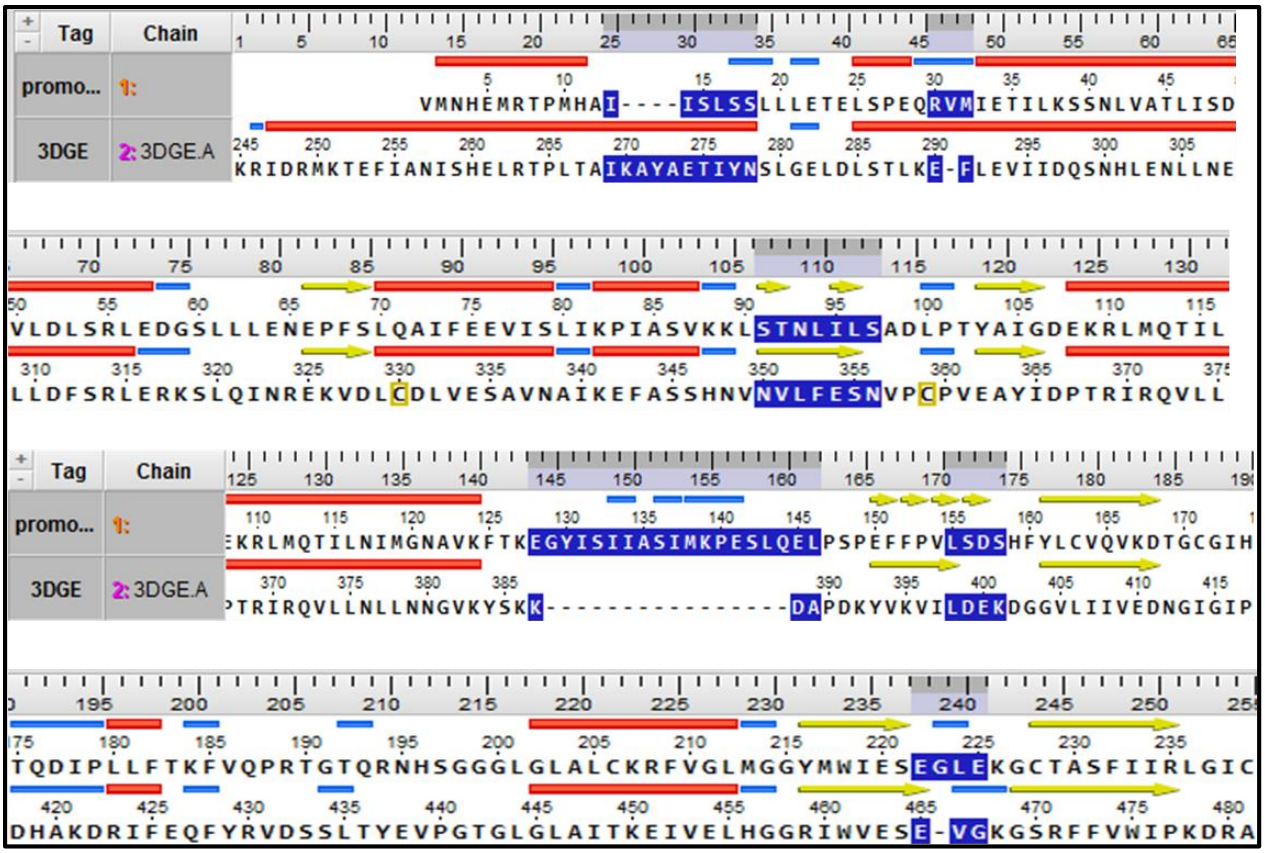



**Figure 28. Structurally diverged regions in target sequence ERS1 kinase domain in comparison to template 3DGE.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel 10 and template as 3DGE.A.

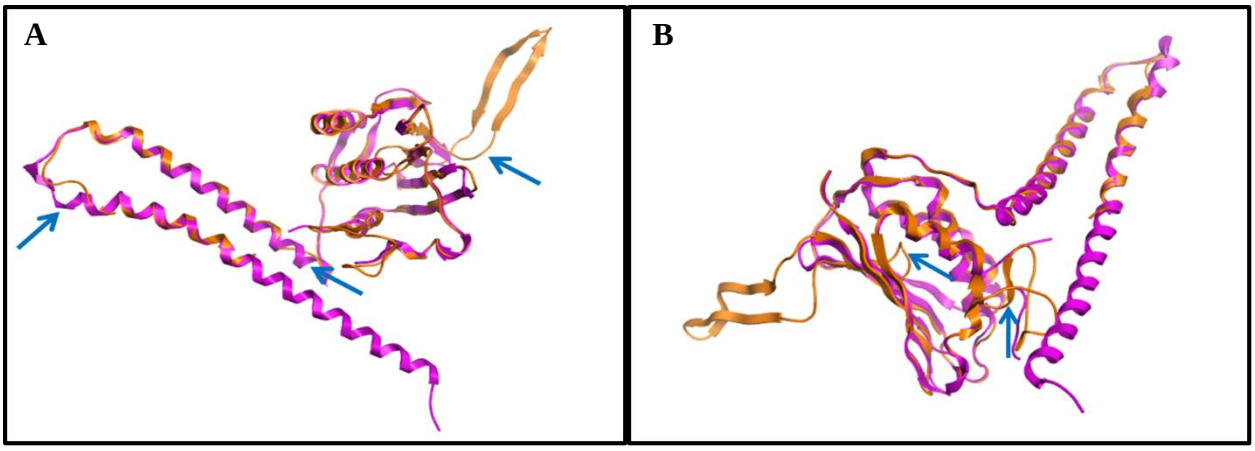



**Figure 29. Three dimensional structure of ETR2 kinase domain using 2C2A. A as template A,B**. Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

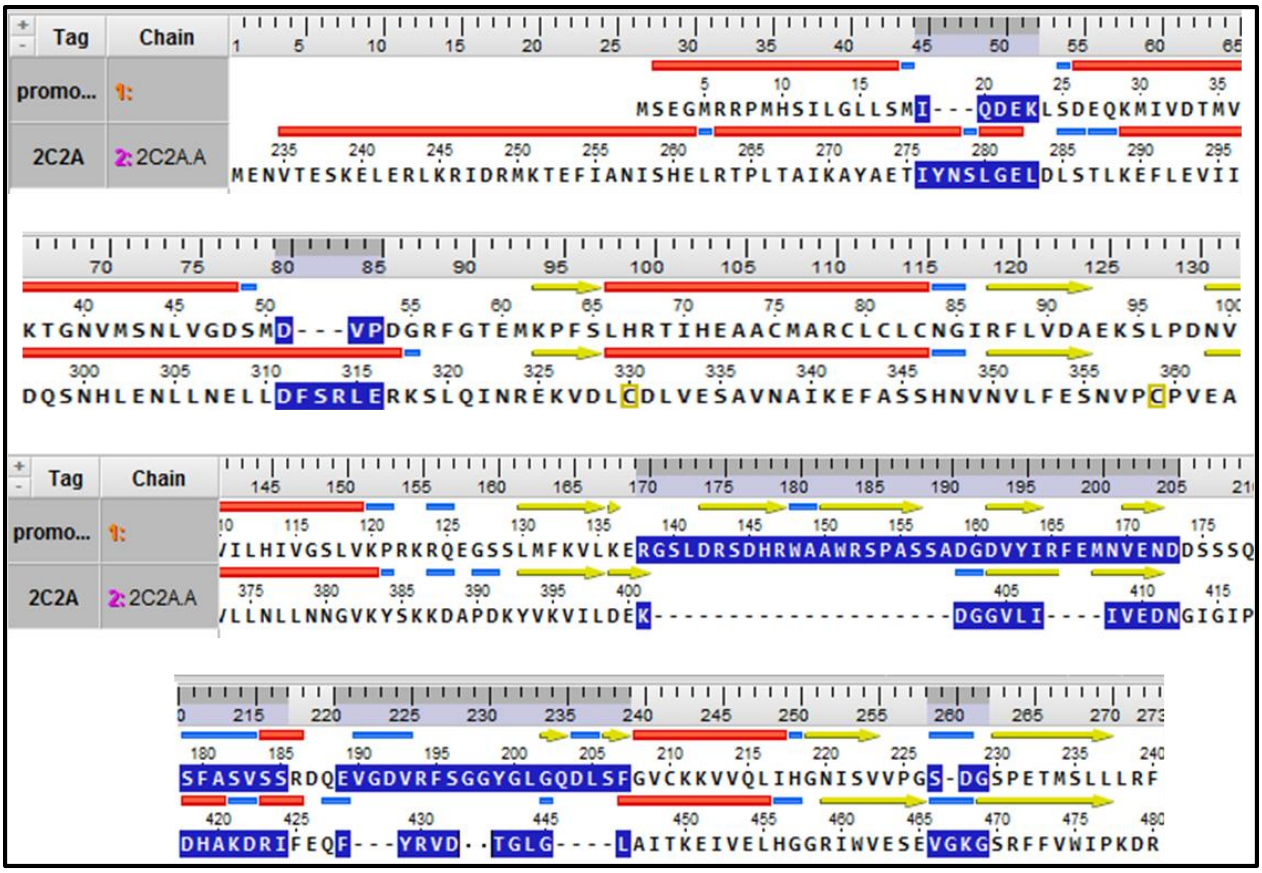



**Figure 30. Structurally diverged regions in target sequence ETR2 kinase domain in comparison to template 2C2A.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel 7 and template as 2C2A.A.

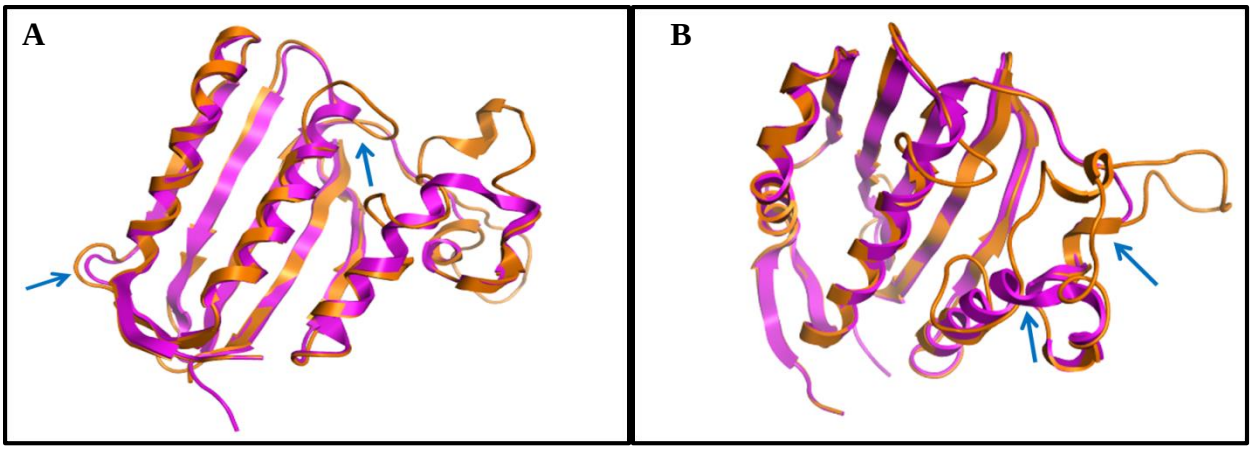



**Figure 31. Three dimensional structure of ETR2 kinase domain using 3A0Y. B as template A,B.** Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

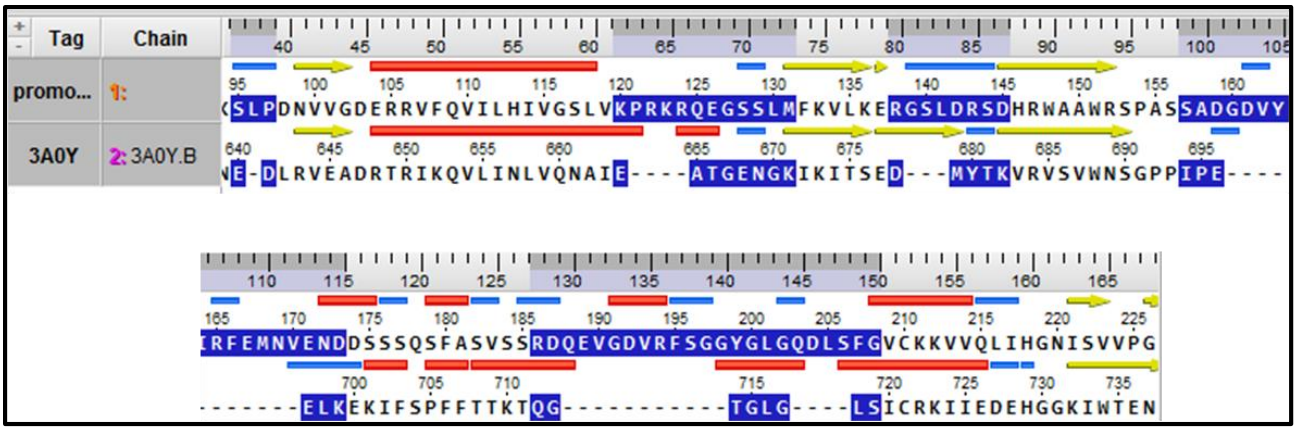



**Figure 32. Structurally diverged regions in target sequence ETR2 kinase domain in comparison to template 3A0Y.B.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel 10 and template as 3A0Y.B.

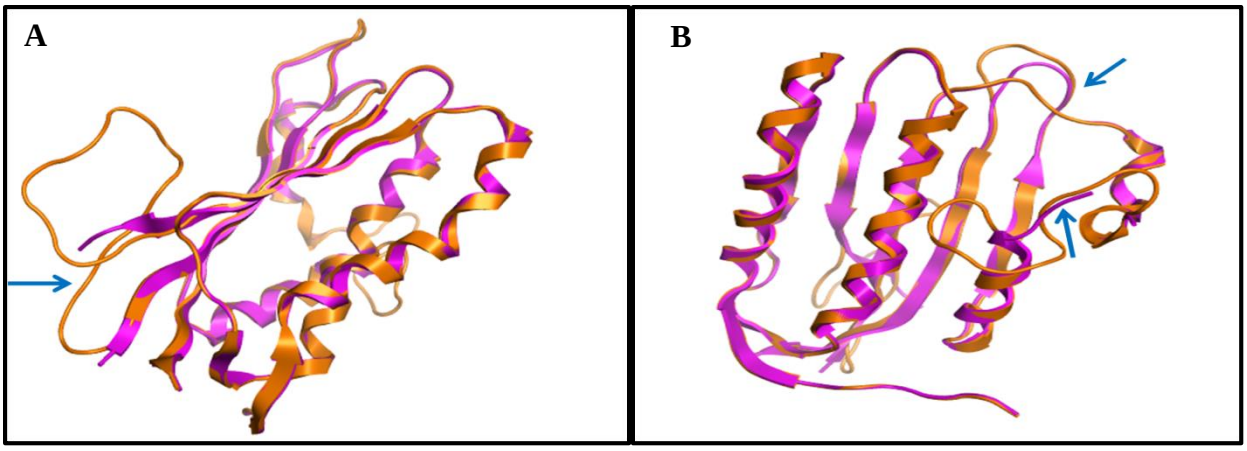



**Figure 33. Three dimensional structure of ERS2 kinase domain using 4PL9. A as template A,B**. Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

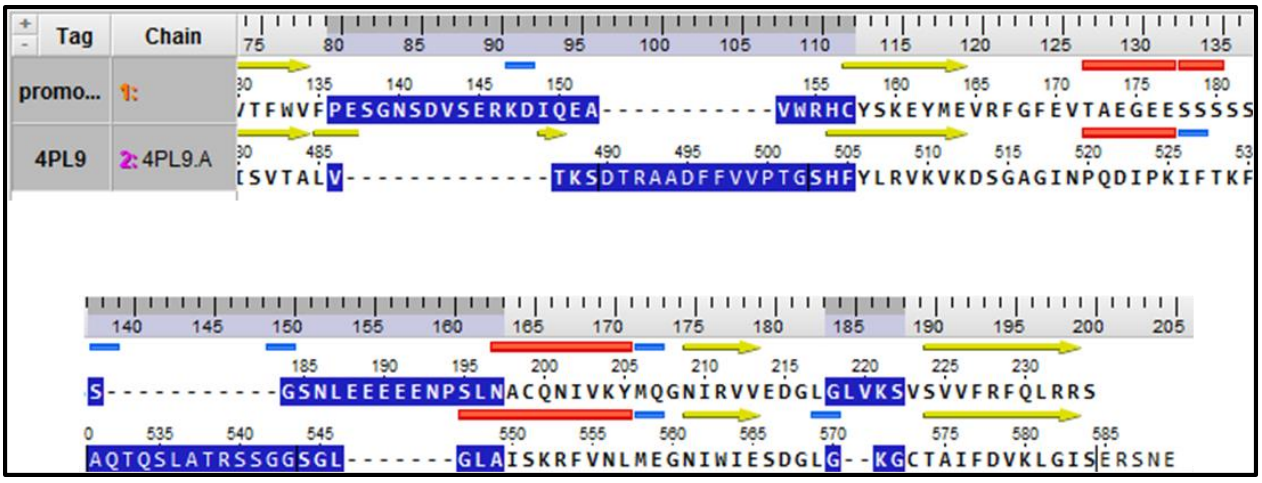



**Figure 34. Structurally diverged regions in target sequence ERS2 kinase domain in comparison to template 4PL9.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel 1 and template as 4PL9.A.

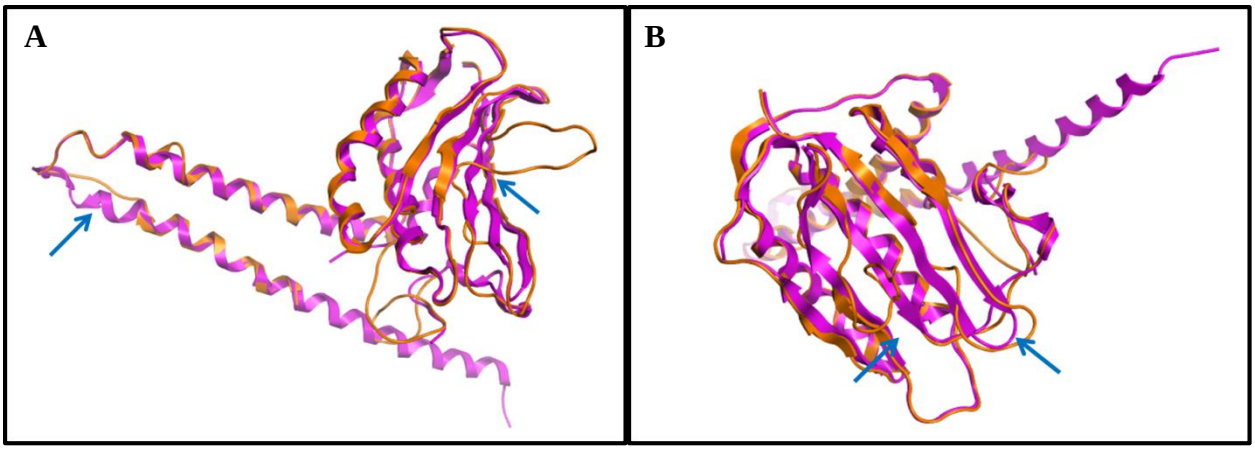



**Figure 35. Three dimensional structure of EIN4 kinase domain using 2C2A. A as template A,B.** Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

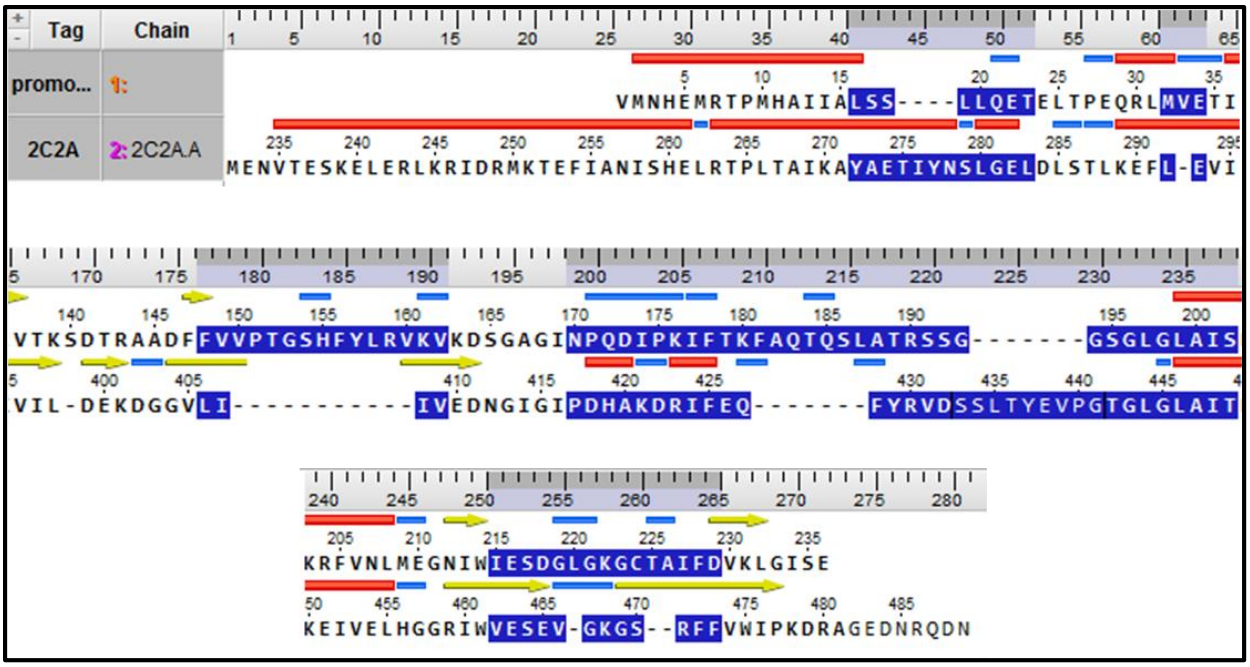



**Figure 36. Structurally diverged regions in target sequence EIN4 kinase domain in comparison to template 2C2A.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel 4 and template as 2C2A.A.

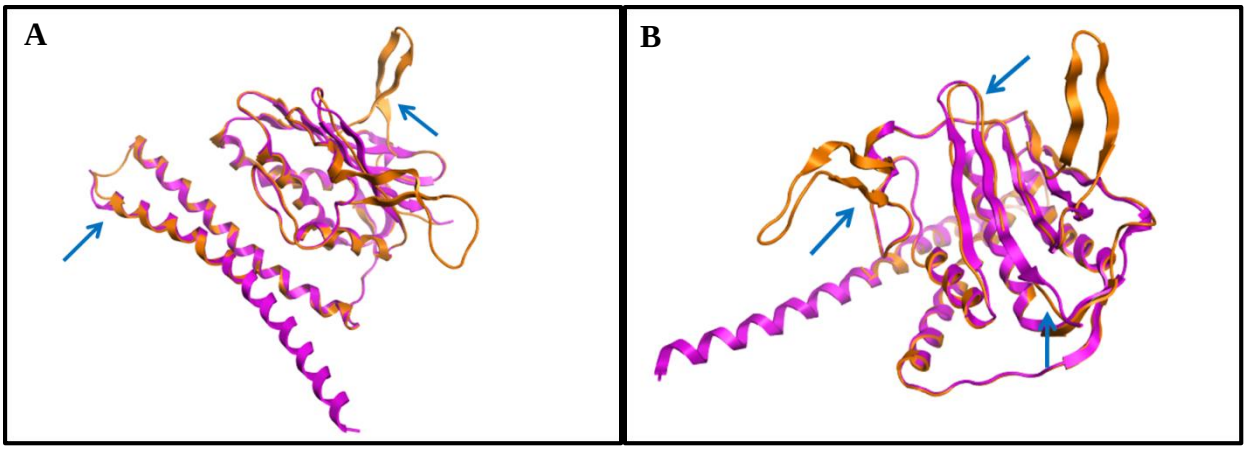



**Figure 37. Three dimensional structure of EIN4 kinase domain using 4Q20.A as template A,B.** Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

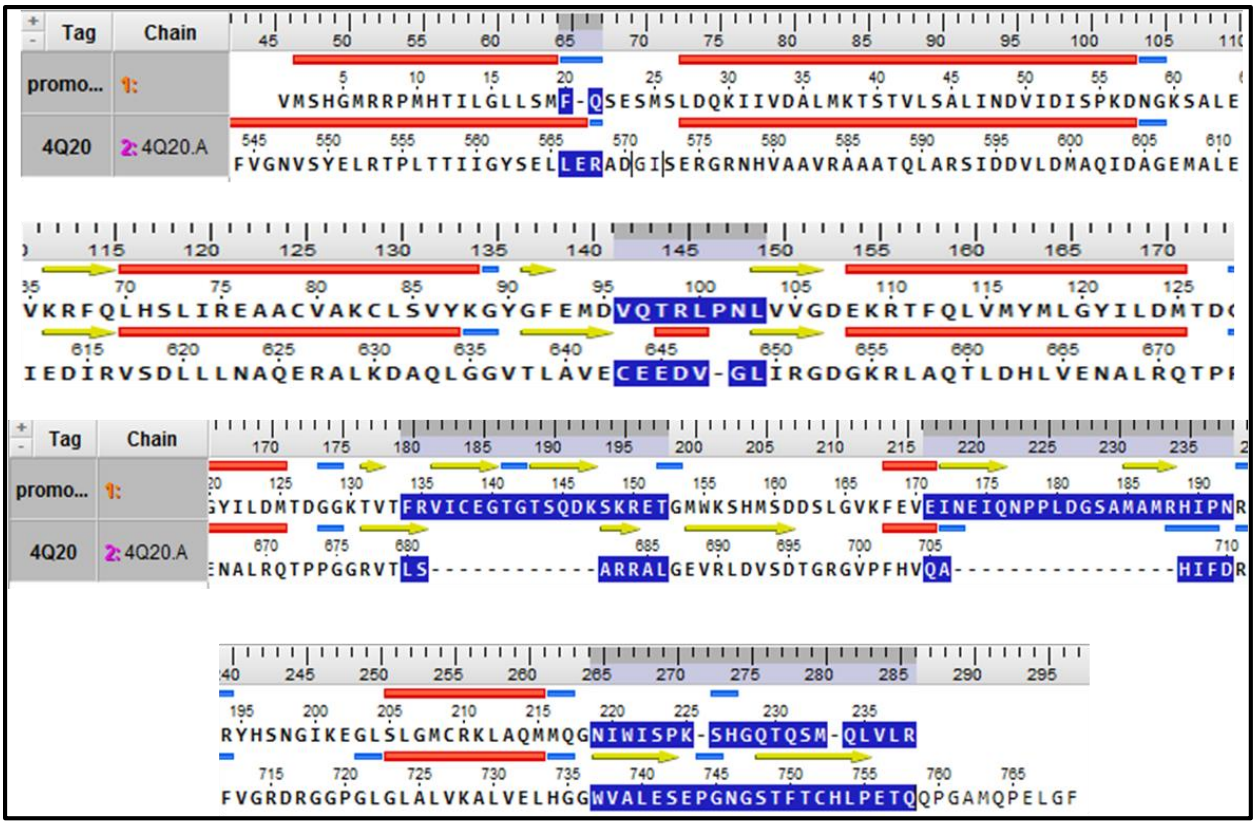



**Figure 38. Structurally diverged regions in target sequence EIN4 kinase domain in comparison to template 4Q20.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel 3 and template as 4Q20.A.

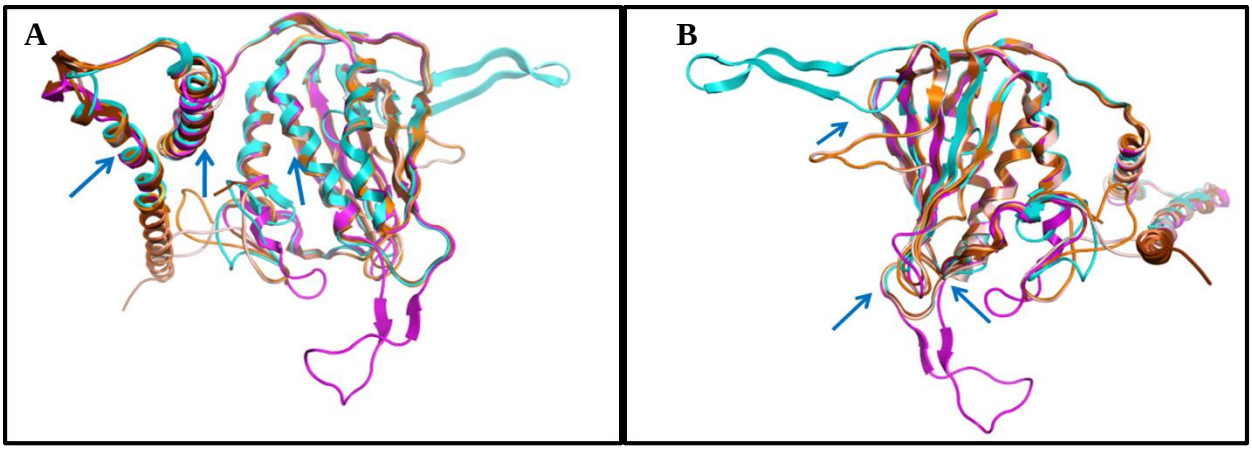



**Figure 39. Comparison of models generated against 2C2A. A for all receptors except ERS2.** The models with least RMSD value generated against 2C2A. A were superimposed in MOE. Panel A represents structurally conserved alpha helices in DHp and CA subunits of kinase domain. Panel B represents structurally diverge regions mainly constituting beta strands of CA subunit. In both the panels, ETR1 was represented in orange, ERS1 in purple, ETR2 in cyan, EIN4 in light brown and crystal structure of the template, 2C2A. A in dark brown colors.

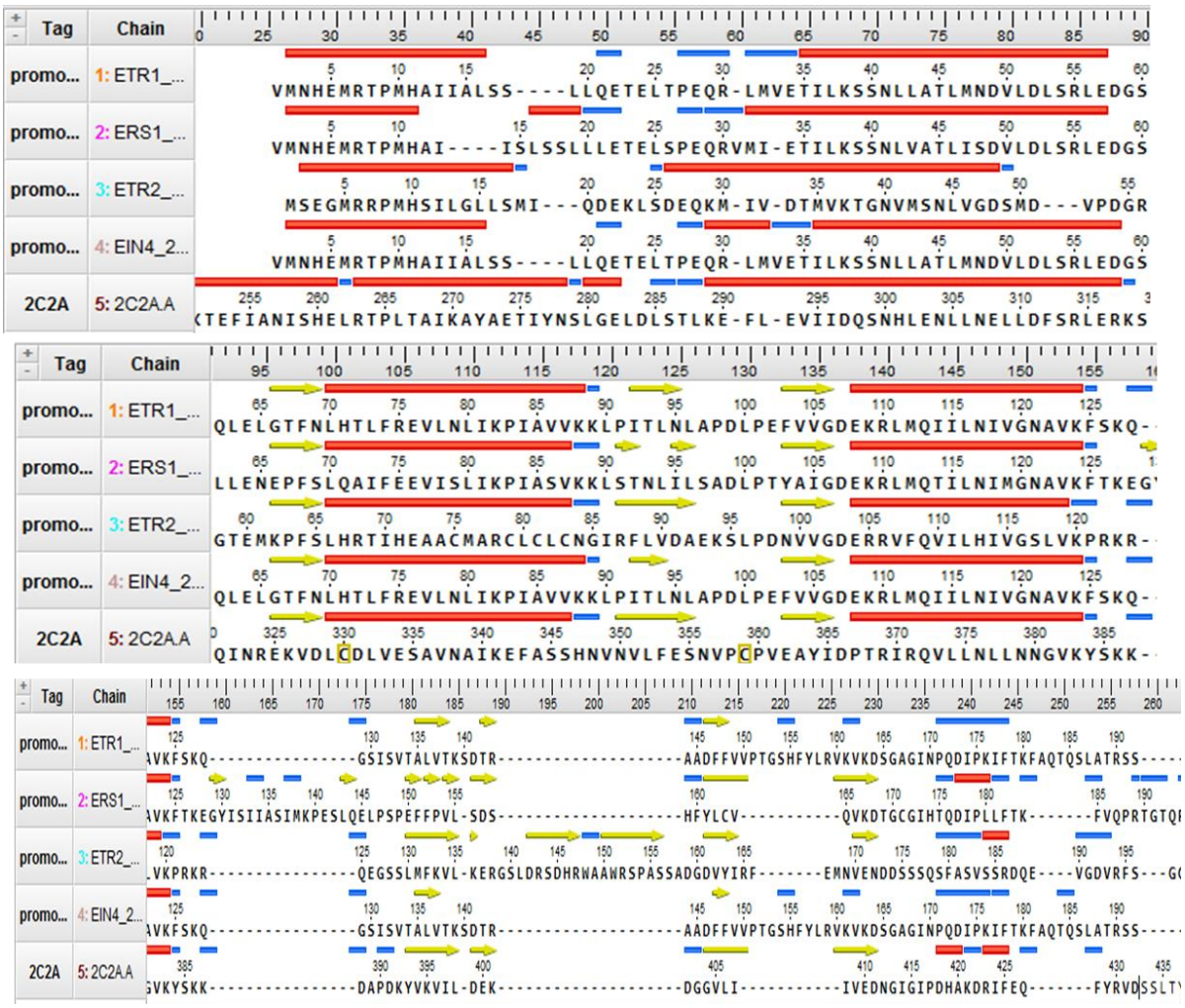
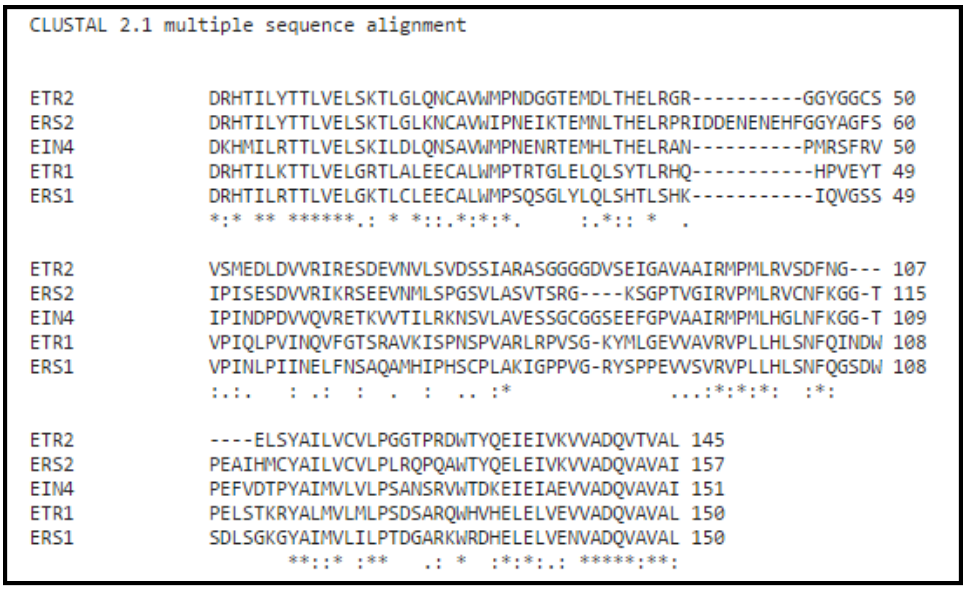



**Figure 40. Structurally diverged/similar regions in target sequences in comparison to template.** The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged/similar regions were selected and represented in sequence Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodels for all the receptors except ERS2 and template as 2C2N. A. The top most panel represents gap in DHp regions corresponding to loop formation in the secondary structure. The middle panel represents the most conserved region among all the receptors. The lower panel represents the most diverge regions on the receptors in comparison to each other as well as with the template.



```
CLUSTAL 2.1 multiple sequence alignment


ETR2            DRHTILYTTLVELSKTLGLQNCAVWMPNDGGTEMDLTHELRGR----------GGYGGCS 50
ERS2            DRHTILYTTLVELSKTLGLKNCAVWIPNEIKTEMNLTHELRPRIDDENENEHFGGYAGFS 60
EIN4            DKHMILRTTLVELSKILDLQNSAVWMPNENRTEMHLTHELRAN----------PMRSFRV 50
ETR1            DRHTILKTTLVELGRTLALEECALWMPTRTGLELQLSYTLRHQ-----------HPVEYT 49
ERS1            DRHTILRTTLVELGKTLCLEECALWMPSQSGLYLQLSHTLSHK-----------IQVGSS 49
                *:* ** ******.: * *::.*:*:*.     :.*:: *  .

ETR2            VSMEDLDVVRIRESDEVNVLSVDSSIARASGGGGDVSEIGAVAAIRMPMLRVSDFNG--- 107
ERS2            IPISESDVVRIKRSEEVNMLSPGSVLASVTSRG----KSGPTVGIRVPMLRVCNFKGG-T 115
EIN4            IPINDPDVVQVRETKVVTILRKNSVLAVESSGCGGSEEFGPVAAIRMPMLHGLNFKGG-T 109
ETR1            VPIQLPVINQVFGTSRAVKISPNSPVARLRPVSG-KYMLGEVVAVRVPLLHLSNFQINDW 108
ERS1            VPINLPIINELFNSAQAMHIPHSCPLAKIGPPVG-RYSPPEVVSVRVPLLHLSNFQGSDW 108
                :.:.   : .:  :  .  :  .. :*              ...:*:*:*:  :*:

ETR2            ----ELSYAILVCVLPGGTPRDWTYQEIEIVKVVADQVTVAL 145
ERS2            PEAIHMCYAILVCVLPLRQPQAWTYQELEIVKVVADQVAVAI 157
EIN4            PEFVDTPYAIMVLVLPSANSRVWTDKEIEIAEVVADQVAVAI 151
ETR1            PELSTKRYALMVLMLPSDSARQWHVHELELVEVVADQVAVAL 150
ERS1            SDLSGKGYAIMVLILPTDGARKWRDHELELVENVADQVAVAL 150
                      **::* :**   .: *  :*:*:.: *****:**:
```

| **Name of the Receptor** | **% Identity** | | | | |
|---|---|---|---|---|---|
| | ETR1_GAF | ERS1_GAF | ETR2_GAF | ERS2_GAF | EIN4_GAF |
| ETR1_GAF Domain | 100 | 64.00 | 40.56 | 29.73 | 42.28 |
| ERS1_ GAF Domain | | 100 | 39.86 | 39.04 | 38.93 |
| ETR2_GAF Domain | | | 100 | 65.25 | 54.48 |
| ERS2_GAF Domain | | | | 100 | 54.42 |
| EIN4_GAF Domain | | | | | 100 |

**Figure 41. Sequence alignment and identity matrix of GAF domains of Ethylene receptors.**

Sequence alignment of the GAF domain of the three ethylene receptors in *Arabidopsis* using CLUSTAL W2. The residues with an asterisk (*) denote conserved residues, colons (**:**) denote conserved substitutions and periods (**.**) denote semi conserved substitutions (Chenna*et al.*, 2003). **B.** The identity matrix was created using the same program which shows the percentage of conserved residues among the receptors. The receptors belonging to sub family I were represented in orange and those belonging to sub family II in green.

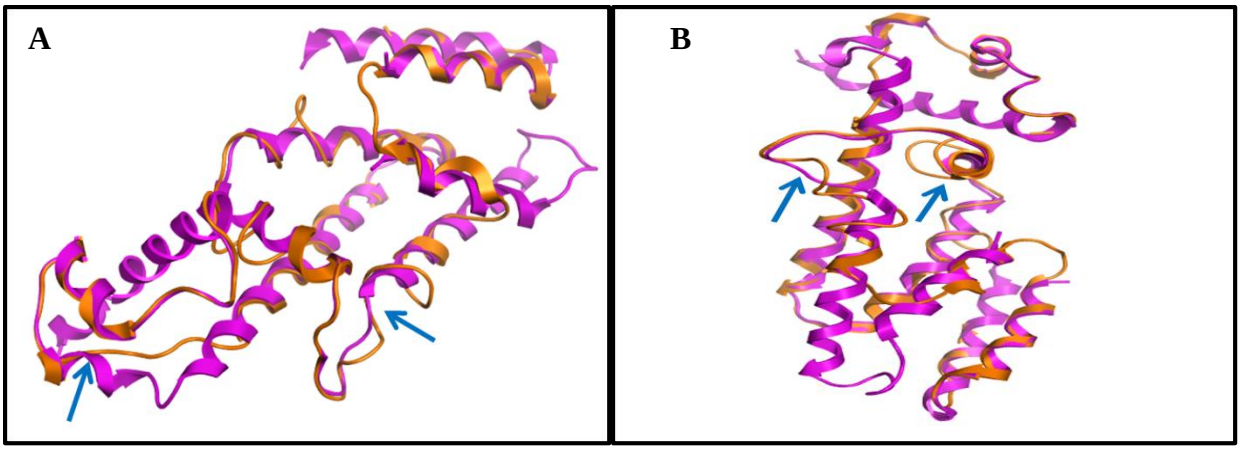



**Figure 42. Three dimensional structure of ETR1 GAF domain using 1UI6.A as template**

**A,B.** Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

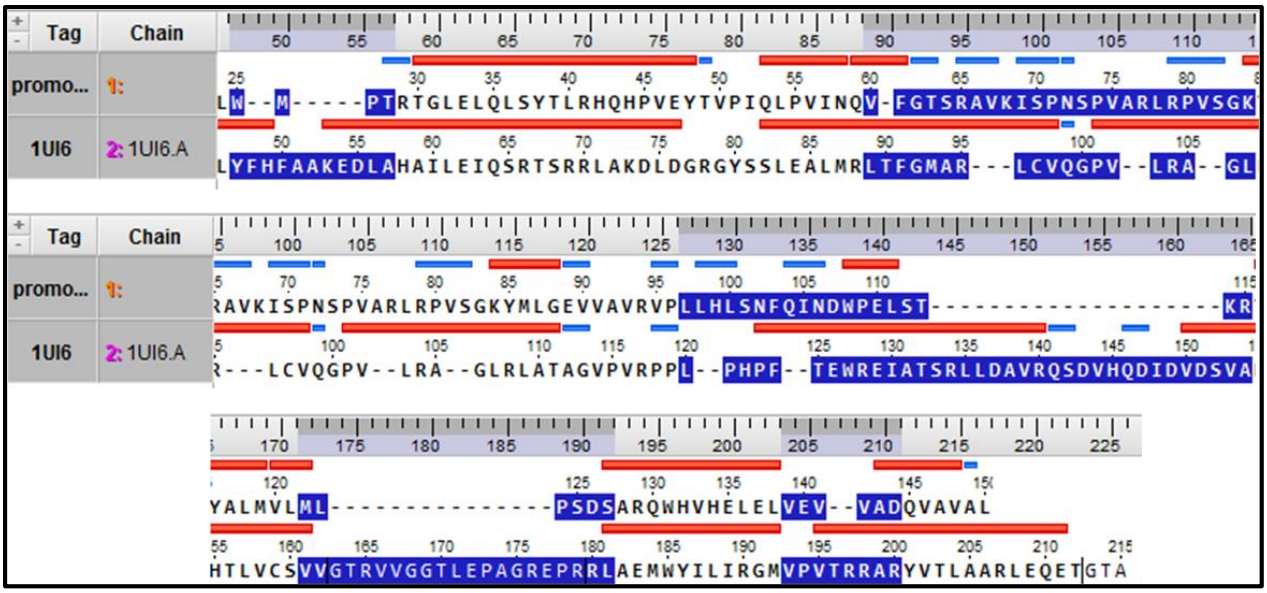



**Figure 43. Structurally diverged regions in target sequence ETR1 GAF domain in comparison to template 1UI6.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel 3 and template as 1UI6.A.

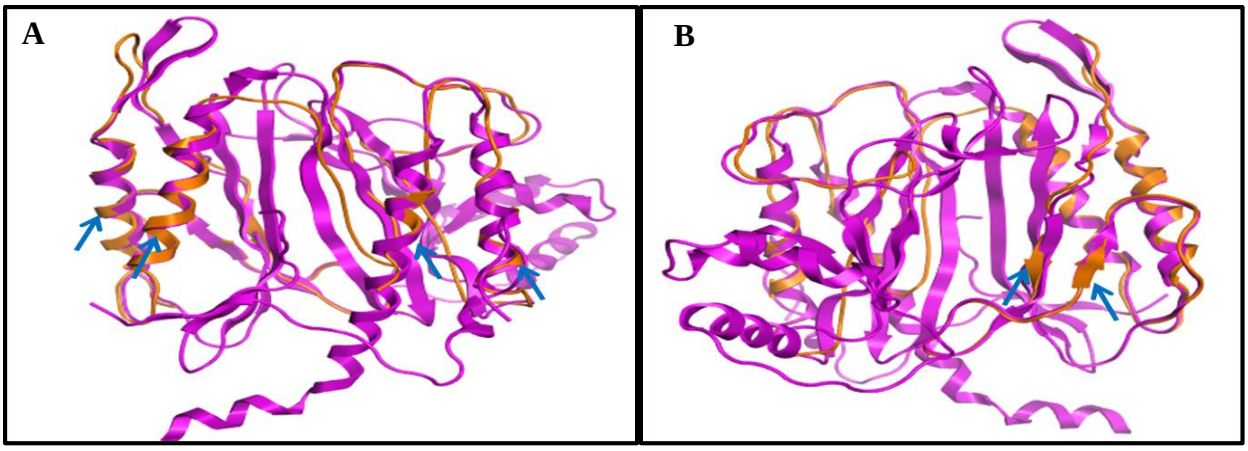



**Figure 44. Three dimensional structure of ETR1 GAF domain using 1B6A.A as template**

**A,B.** Structurally similar regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

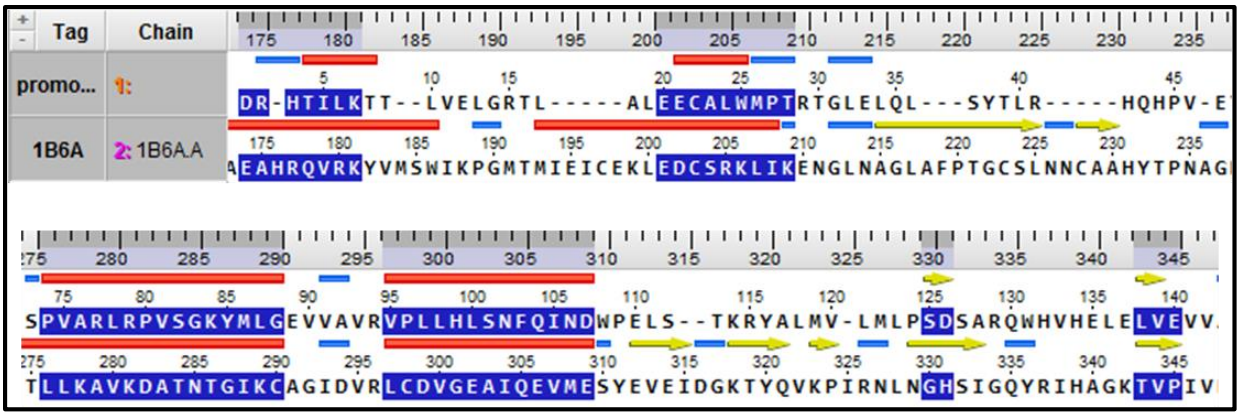



**Figure 45. Structurally similar/conserved regions in target sequence ETR1 GAF domain in comparison to template 1B6A.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally conserved regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel with least RMSD value and template as1B6A.A.

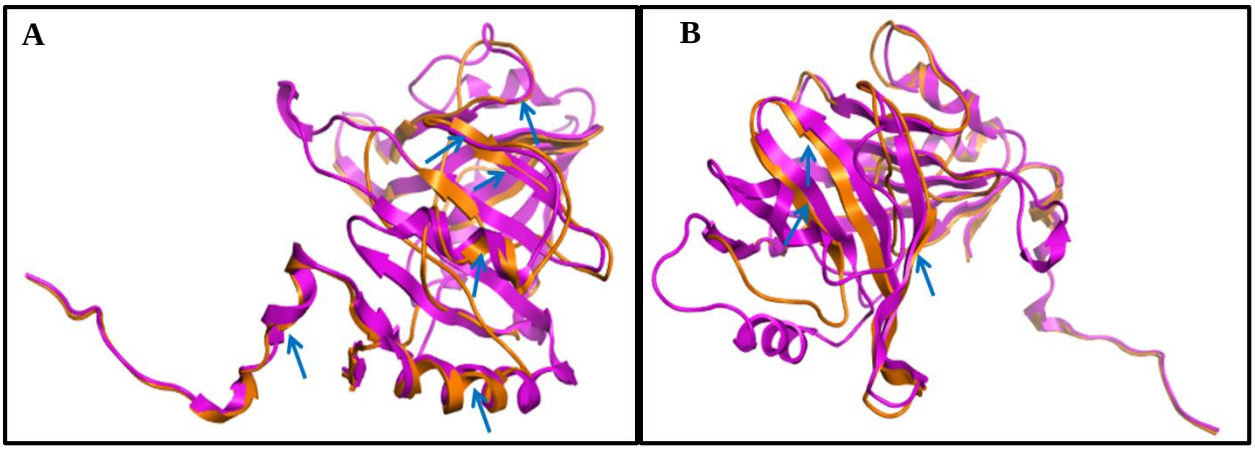



**Figure 46. Three dimensional structure of ERS1 GAF domain using 3TTG. A as template A,B.** Structurally conserved regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

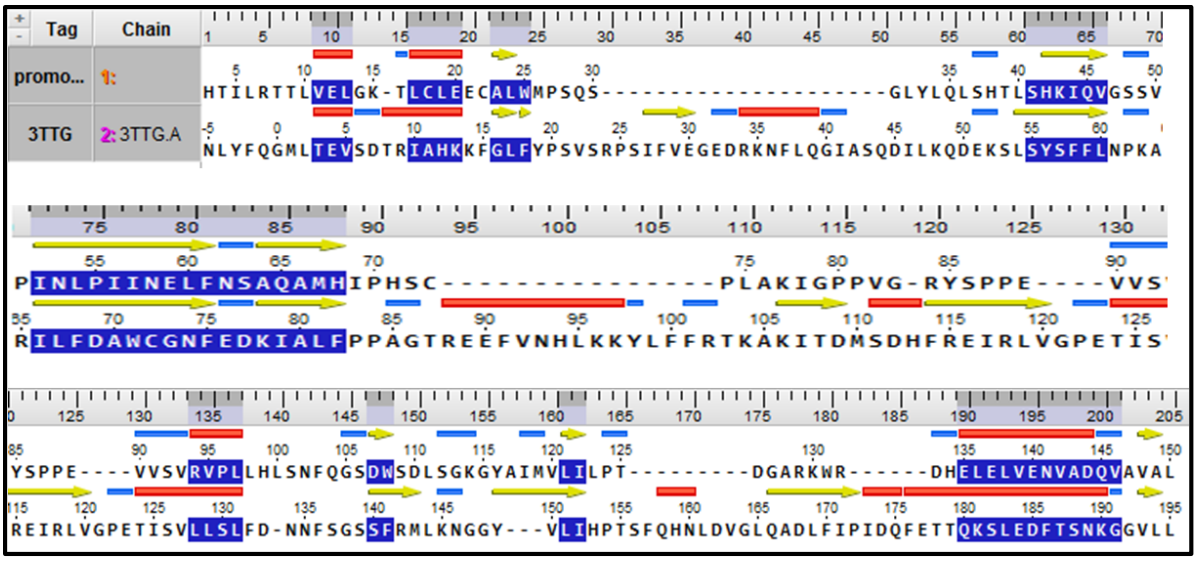



**Figure 47. Structurally similar/conserved regions in target sequence ERS1 GAF domain in comparison to template 3TTG.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally conserved regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel with least RMSD value and template as 3TTG.A.

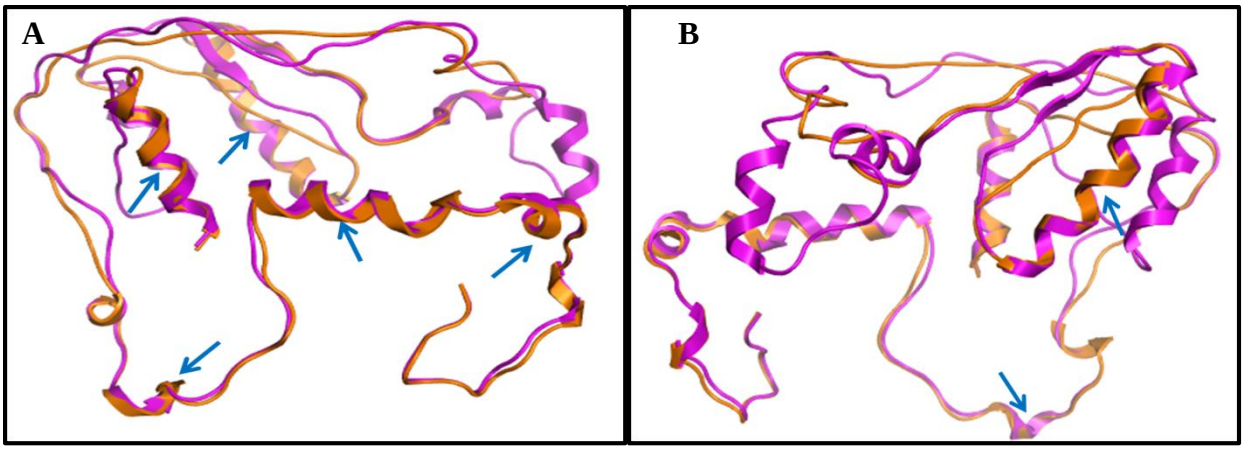



**Figure 48. Three dimensional structure of ERS1 GAF domain using 3FY4.A as template A,B.** Structurally conserved regions in the model that were conformationally similar from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

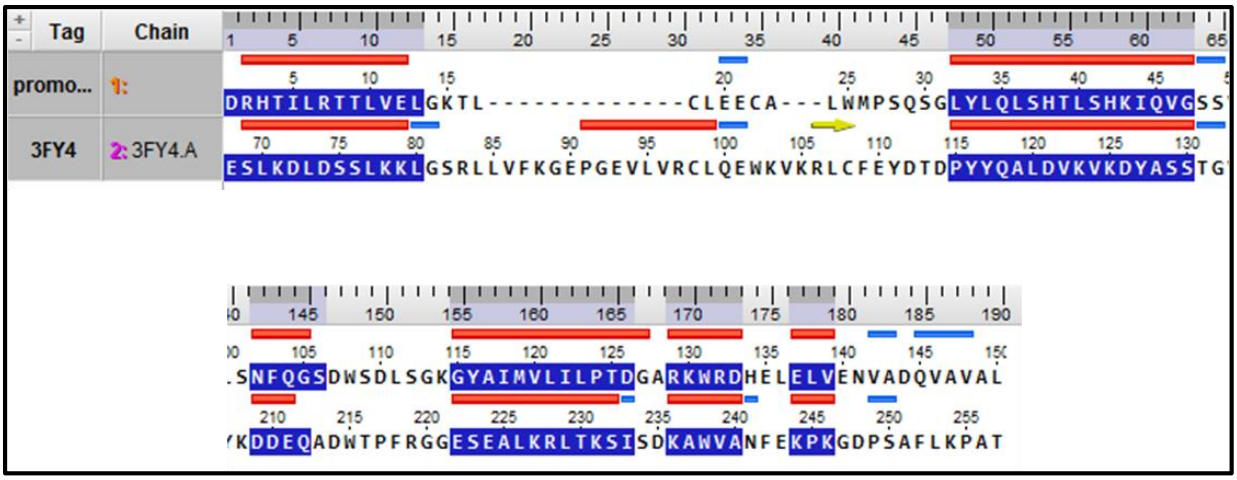



**Figure 49. Structurally similar/conserved regions in target sequence ERS1 GAF domain in comparison to template, 3FY4.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally conserved regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel with least RMSD value and template as 3FY4.A.

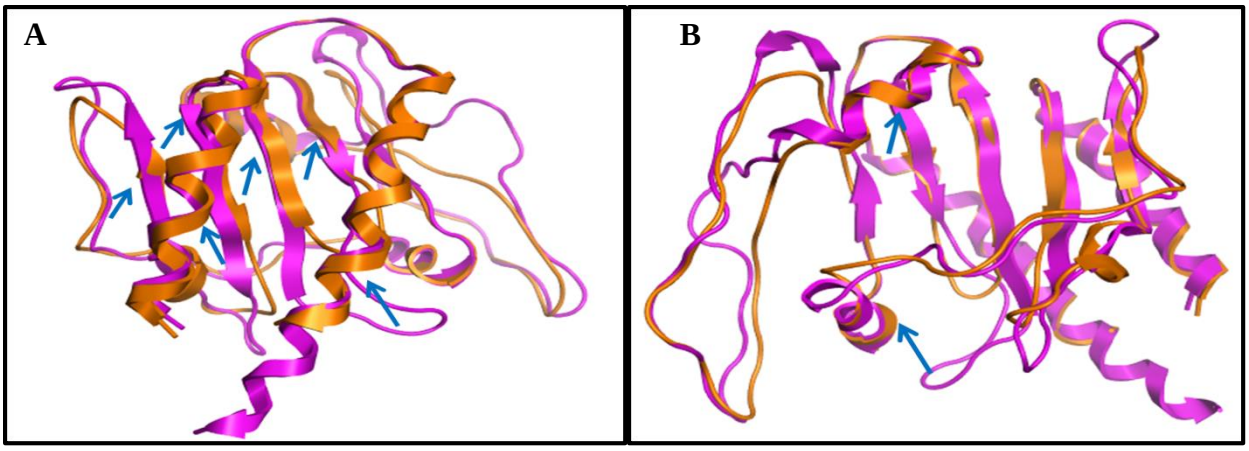



**Figure 50. Three dimensional structure of ETR2 GAF domain using 2K2N.A as template A,B.** Structurally conserved regions in the model that were conformationally similar from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

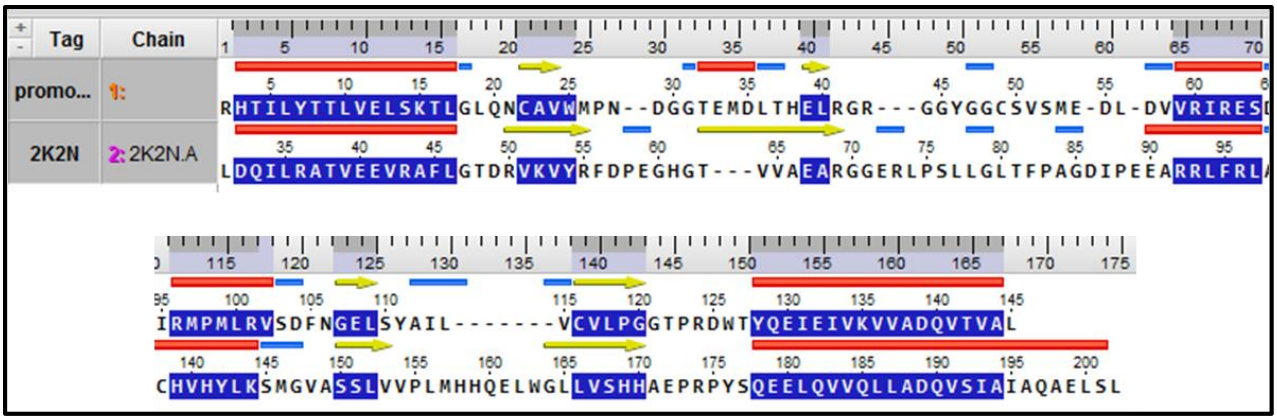



**Figure 51. Structurally similar/conserved regions in target sequence ETR2 GAF domain in comparison to template 2K2N.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally conserved regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel with least RMSD value and template as 2K2N.A.

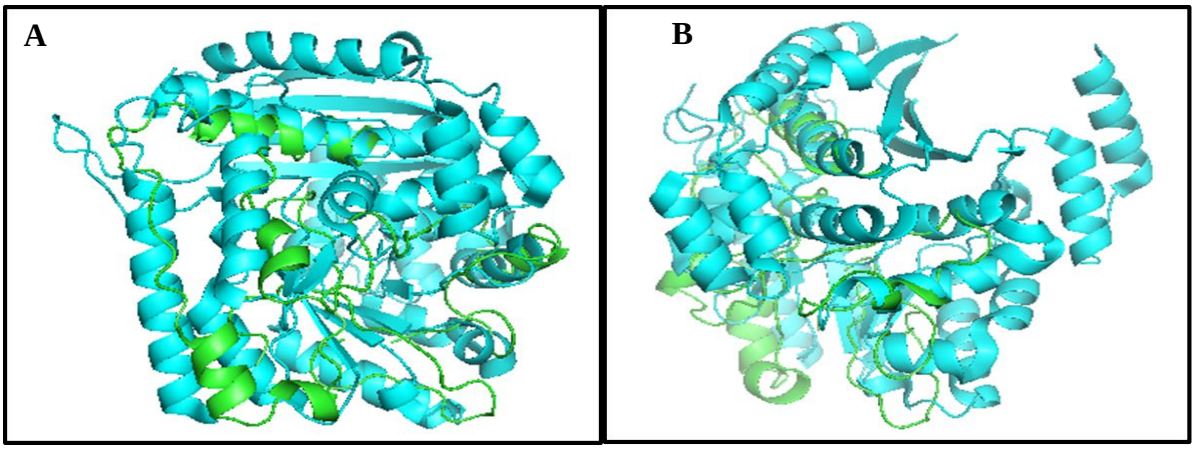



**Figure 52. Three dimensional structure of ETR2 GAF domain using 4HL7.A as template**

**A, B.** Superimposing least diverged model with the crystal structure and representation of basic secondary structure generated for ETR2 GAF domain superimposed on 4HL7.A using 'align' function in PyMol. In each panel, the model was represented in green and the crystal structure was represented in cyan.

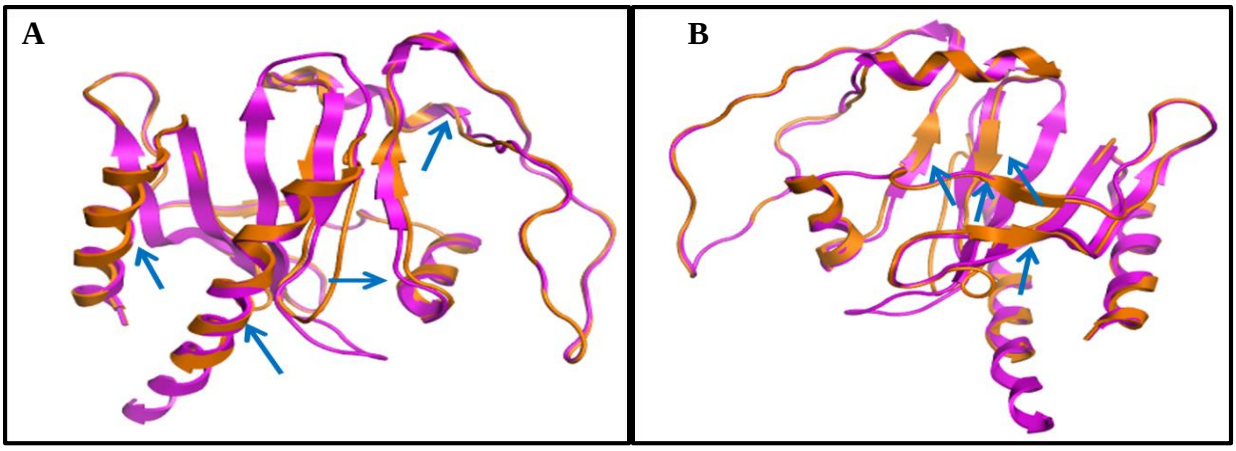



**Figure 53. Three dimensional structure of ERS2 GAF domain using 2K2N. A as template A,B.** Structurally conserved regions in the model that were conformationally similar to respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

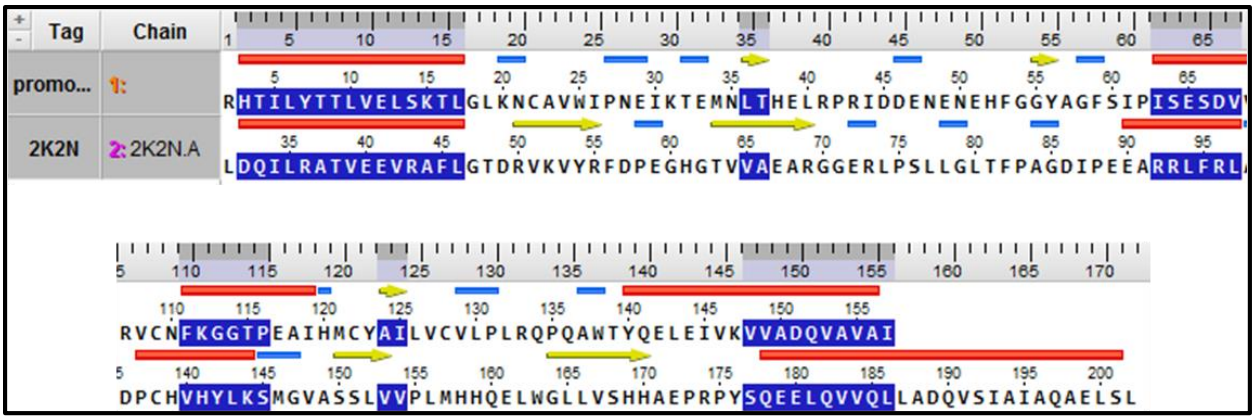



**Figure 54. Structurally similar/conserved regions in target sequence ERS2 GAF domain in comparison to template 2K2N.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally conserved regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel with least RMSD value and template as 2K2N.A.

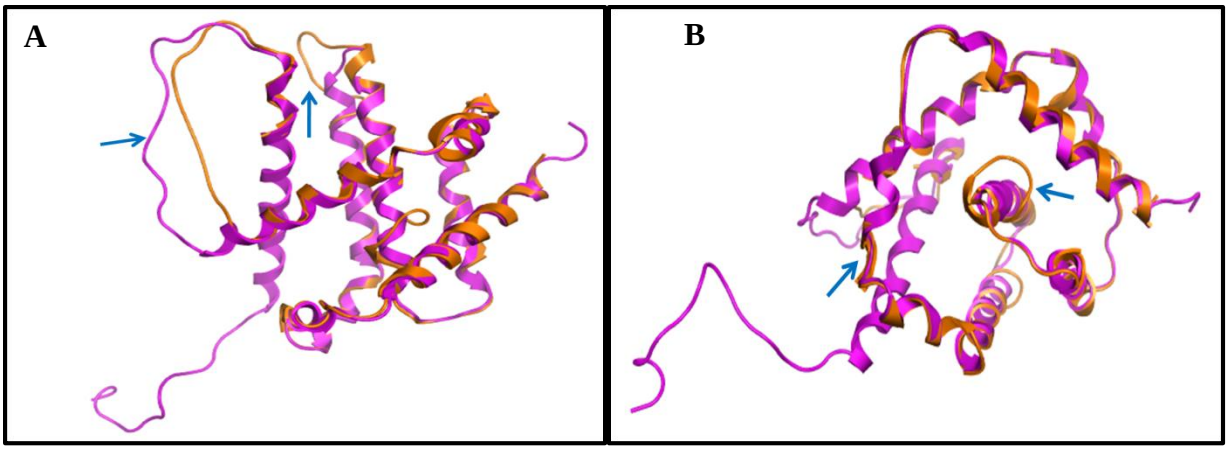



**Figure 55. Three dimensional structure of ERS2 GAF domain using 2K7W.A as template A,B.** Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

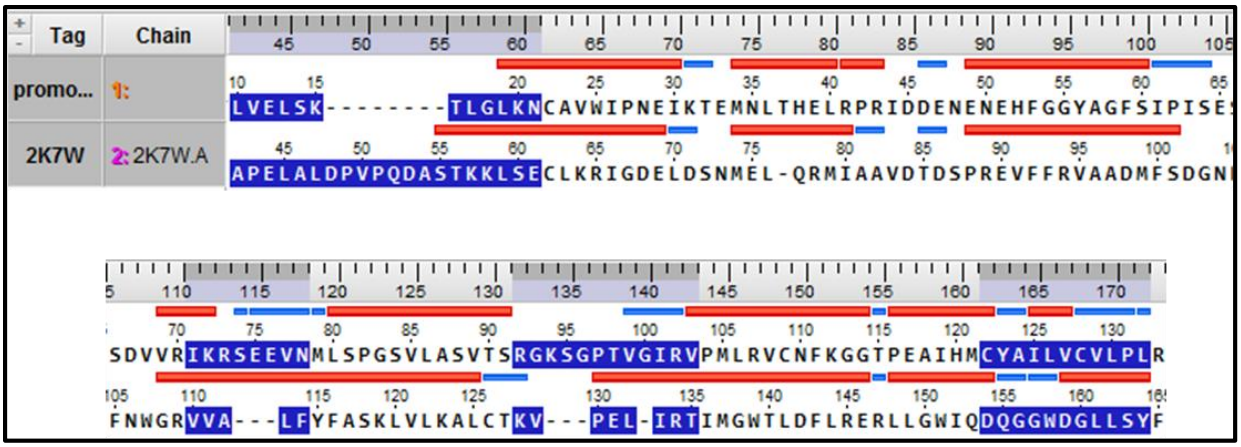



**Figure 56. Structurally diverge regions in target sequence ERS2 GAF domain in comparison to template 2K7W.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel with least RMSD value and template as 2K7W.A.

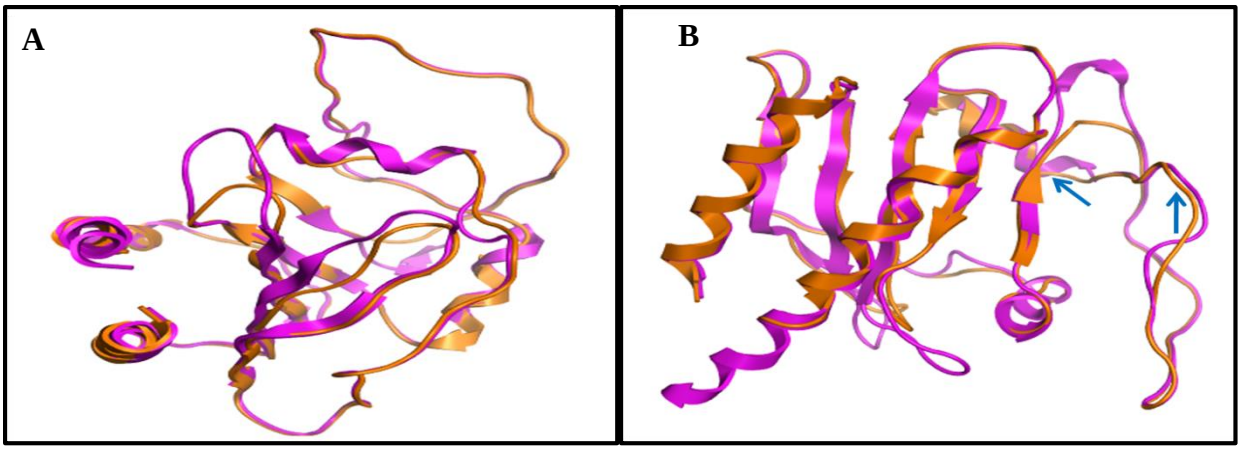



**Figure 57. Three dimensional structure of EIN4 GAF domain using 2K2N.A as template. A,B**. Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

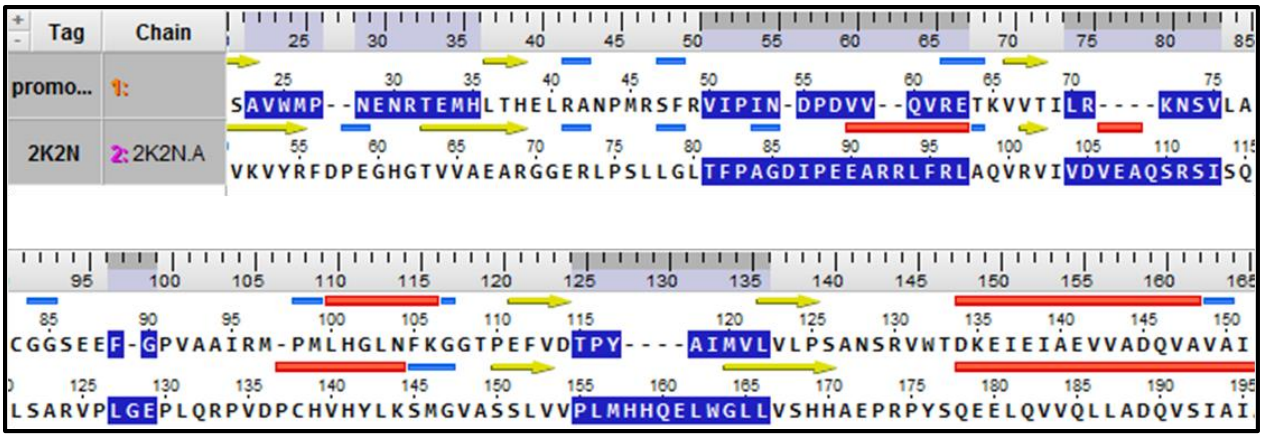



**Figure 58. Structurally diverge regions in target sequence EIN4 GAF domain in comparison to template 2K2N.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel with least RMSD value and template as 2K2N.A.

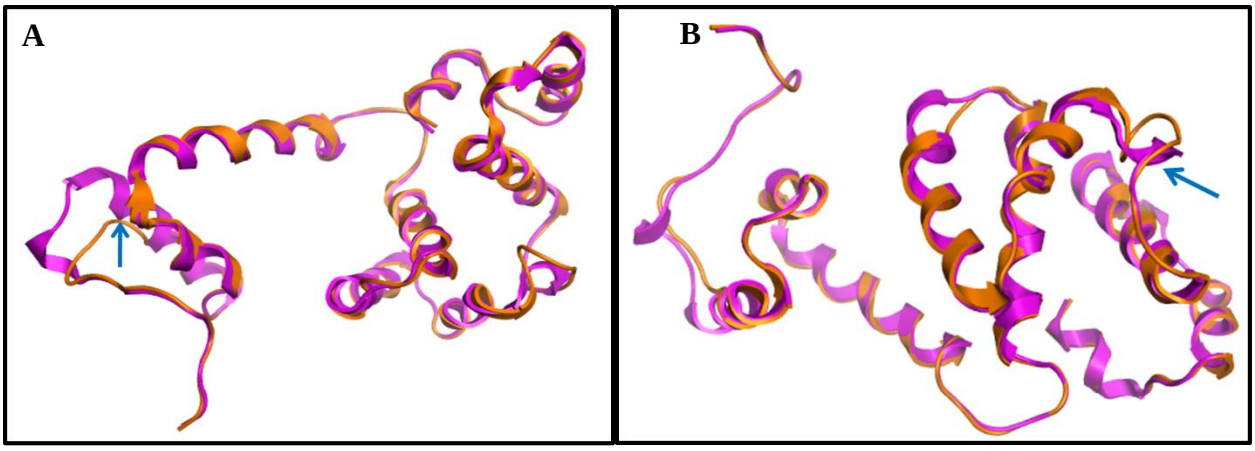



**Figure 59. Three dimensional structure of EIN4 GAF domain using 2PRR.A as template A,B.** Structurally diverged regions in the model that were conformationally different from respective regions in the template were indicated in bold blue arrow. In each panel, the model was represented in orange and the crystal structure in purple.

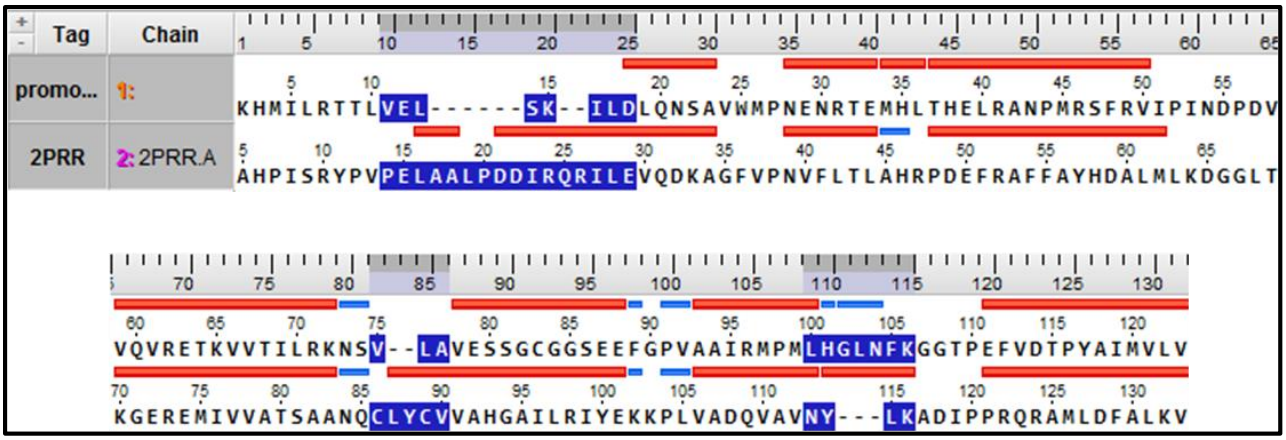



**Figure 60. Structurally diverge regions in target sequence EIN4 GAF domain in comparison to template 2PRR.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequence was denoted as Promodel with least RMSD value and template as 2PRR.A.

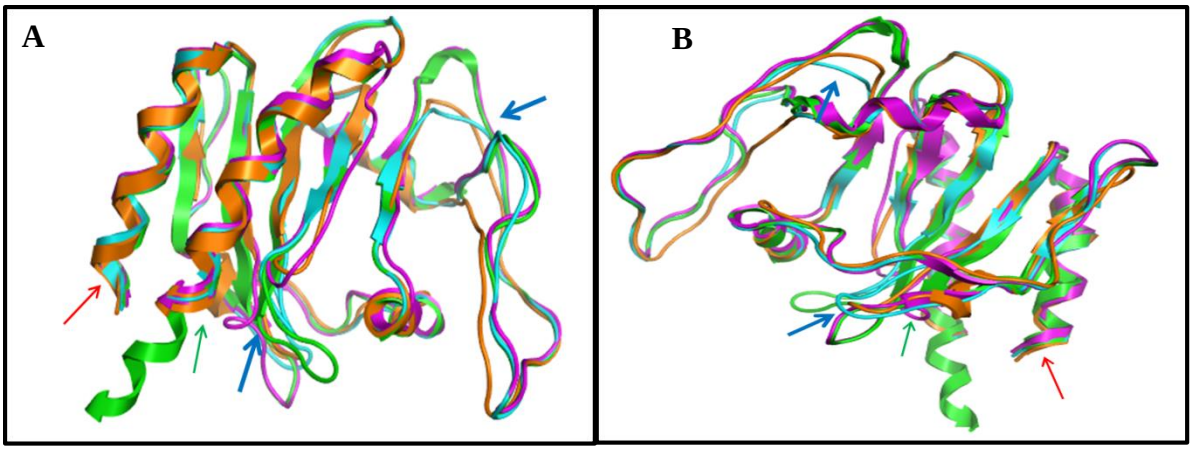



**Figure 61. Superimposing three dimensional structures of subfamily II GAF domains using 2K2N.A as template**

**A,B.** Superimposing least diverged models and representation of possible different orientations of all the models generated for sub family II GAF domain superimposed on 2K2N.A using 'superimpose' function in MOE. Regions that were conformationally different in the model to their respective regions in the template were indicated with bold blue arrows. In each panel, the model for ETR2 was represented in orange, ERS2 in purple, EIN4 in Cyan and the crystal structure of 2K2N.A was represented in green color.

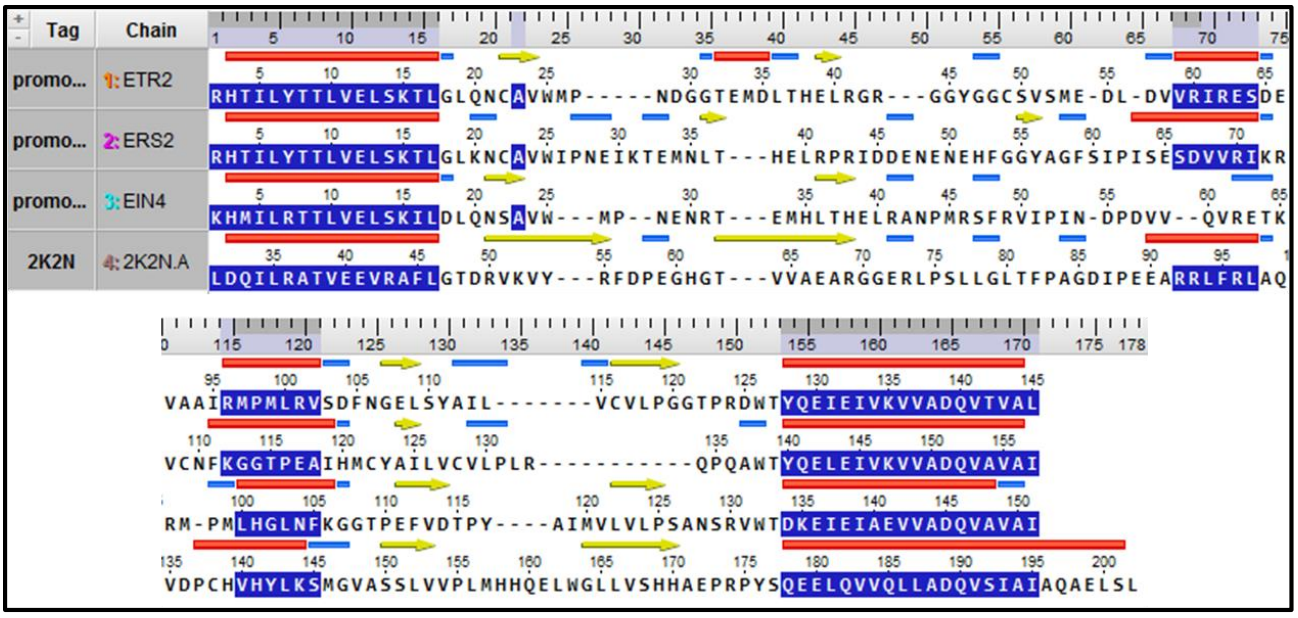



**Figure 62. Structurally similar regions in target sequences of GAF domains of subfamily II receptors in comparison to template 2K2N.A.**

The secondary structures were assigned to the sequences in MOE and atoms with structurally diverged regions were selected and represented in sequence as highlighted in blue. Alpha helices were indicated as longitudinal red bars, beta strands were represented as yellow arrows, and bends as short blue lines. Target sequences were denoted as Promodels with least RMSD value of respective receptors and template as 2K2N.A.

# Tables

**Table 1. List of different templates selected by MOE based on Z scores from PDB, for ETR2 receiver domain protein sequence using search function in MOE.**

The table provides information on the full names of the templates along with their PDB ID and Z scores with ETR2 sequence. The chain name used as the template was indicated after the period in the PDB ID.

| SNo. | Z Score | PDB ID | Name of the Template |
|---|---|---|---|
| 1. | Good | 1DCF.A | Crystal Structure Of The Receiver Domain Of The Ethylene Receptor Of Arabidopsis Thaliana |
| 2. | 20.7 | 2AYX.A | Solution Structure Of The E.ColiRcsc C-Terminus (Residues 700-949) Containing Linker Region And Phosphoreceiver Domain |
| 3. | 13.9 | 3R0J.A | Structure Of Phop From Mycobacterium Tuberculosis |
| 4. | 12.4 | 3RVK.A | Structure Of The Chey-Mn2+ Complex With Substitutions At 59 And 89: N59d E89q |
| 5. | 11.8 | 3C3M.A | Crystal Structure Of The N-Terminal Domain Of Response Regulator Receiver Protein From MethanoculleusMarisnigri Jr1 |
| 6. | 12.6 | 3EQ2.A | Structure Of Hexagonal Crystal Form Of Pseudomonas Aeruginosa Rssb |
| 7. | 12.7 | 3CG0.C | Crystal Structure Of Signal Receiver Domain Of Modulated Diguanylate Cyclase From DesulfovibrioDesulfuricans G20, An Example Of Alternate Folding |
| 8. | 10.3 | 1W25.A | Response Regulator Pled In Complex With C-Digmp |
| 9. | 9.3 | 3C97.A | Crystal Structure Of The Response Regulator Receiver Domain Of A Signal Transduction Histidine Kinase From AspergillusOryzae |
| 10. | 9.3 | 1P2F.A | Crystal Structure Analysis Of Response Regulator Drrb, A ThermotogaMaritimaOmpr/Phob Homolog |
| 11. | 9.9 | 2HQR.A | Structure Of A Atypical Orphan Response Regulator Protein Revealed A New Phosphorylation-Independent Regulatory Mechanism |

**Table 2. List of different templates selected by MOE based on Z scores from PDB, for EIN4receiver domain protein sequence using search function in MOE.**

The table provides information on the full names of the templates along with their PDB ID and Z scores with EIN4 sequence. The chain name used was the template is indicated after the period in the PDB ID.

| SNo. | Z Score | PDB ID | Name of the Template |
|---|---|---|---|
| 1. | Good | 1DCF.A | Crystal Structure Of The Receiver Domain Of The Ethylene Receptor Of Arabidopsis thaliana |
| 2. | Good | 2AYX.A | Solution Structure Of The E.ColiRcsc C-Terminus (Residues 700-949) Containing Linker Region And PhosphoreceiverDomain |
| 3. | 19.5 | 3C3M.A | Crystal Structure Of The N-Terminal Domain Of Response Regulator Receiver Protein From Methanoculleusmarisnigri Jr1 |
| 4. | 19.0 | 1W25.A | Response Regulator Pled In Complex With C-Digmp |
| 5. | 17.3 | 3EQ2.A | Structure Of Hexagonal Crystal Form Of Pseudomonas Aeruginosa Rssb |
| 6. | 16.3 | 3R0J.A | Structure of PhoP from Mycobacterium tuberculosis |
| 7. | 16.1 | 2R25.B | Complex of YPD1 and SLN1-R1 with bound Mg2+ and BeF3- |
| 8. | 15.4 | 1P2F.A | Crystal Structure Analysis of Response Regulator DrrB, a ThermotogamaritimaOmpR/PhoB Homolog |
| 9. | 14.9 | 2WB4.A | ACTIVATED DIGUANYLATE CYCLASE PLED IN COMPLEX WITH C-DI-GMP |
| 10. | 14.8 | 3BRE.A | Crystal Structure of P.aeruginosa PA3702 |
| 11. | 14.6 | 1YS7.A | Crystal structure of the response regulator protein prrAcomlexed with Mg2+ |
| 12. | 14.3 | 3LUF.B | Structure of probable two-component system response regulator/GGDEF domain protein |
| 13. | 14.0 | 2V0N.B | Activated Response Regulator Pled In Complex With C-Digmp And Gtp-Alpha-S |

**Table 3. List of different templates selected by MOE based on Z scores from PDB, for ETR1 kinase domain protein sequence using search function in MOE and PDB BLAST.**

The table provides information on the full names of the templates along with their PDB ID and Z scores with ETR1 sequence. The chain name used as the template was indicated after the period in the PDB ID.

| SNo. | Z/E Score | PDB ID | Name of the Template |
|---|---|---|---|
| 1. | Good | 2C2A.A | Structure of the entire cytoplasmic portion of a sensor histidine kinase protein |
| 2. | Good | 3DGE.A | Structure of a histidine kinase-response regulator complex reveals insights into two-component signaling and a novel cis-autophosphorylation mechanism |
| 3. | Good | 4A2L.E | Structure of the periplasmic domain of the heparin and heparan sulphate sensing hybrid two component system bt4663 in apo and ligand bound forms |
| 4. | 17.1 | 3A0Y.B | Catalytic domain of histidine kinase thka (tm1359) (nucleotide free form 3: 1,2-propanediol, orthorombic) |
| 5. | 16.2 | 3JZ3.A | Structure of the cytoplasmic segment of histidine kinase qsec |
| 6. | 15.4 | 1BXD.A | NMR structure of the histidine kinase domain of the *e. coli* osmosensor envz |
| 7. | 15.5 | 3SL2.A | ATP forms a stable complex with the essential histidine kinase walk (yycg) domain |
| 8. | 11.0 | 3D36.A | Crystal structure of *Geobacillus stearothermophilus* kinb with the inhibitor sda |
| 9. | 10.1 | 4EW8.A | Crystal structure of a c-terminal part of tyrosine kinase (divl) from *Caulobacter crescentus* cb15 at 2.50å resolution (psi community target, shapiro) |
| 10. | 7.2 | 1ID0.A | Crystal structure of the nucleotide bond conformation of PHOQ kinase domain |
| 11. | 3e-116 | 4PL9.A | Structure of the catalytic domain of ETR1 from *Arabidopsis thaliana* |

**Table 4. List of different templates selected by MOE based on Z scores from PDB, for ERS1 kinase domain protein sequence using search function in MOE and PDB BLAST.**

The table provides information on the full names of the templates along with their PDB ID and Z scores with ERS1 sequence. The chain name used as the template was indicated after the period in the PDB ID.

| SNo. | Z/E Score | PDB ID | Name of the Template |
|---|---|---|---|
| 1. | Good | 3DGE.A | Structure of a histidine kinase-response regulator complex reveals insights into two-component signaling and a novel cis-autophosphorylation mechanism |
| 2. | 23.0 | 2C2A.A | Structure of the entire cytoplasmic portion of a sensor histidine kinase protein |
| 3. | 16.0 | 3JZ3.A | Structure of the cytoplasmic segment of histidine kinase qsec |
| 4. | 17.4 | 3A0Y.B | Catalytic domain of histidine kinase thka (tm1359) (nucleotide free form 3: 1,2-propanediol, orthorombic) |
| 5. | 13.4 | 3D36.A | Crystal structure of *Geobacillus stearothermophilus* kinb with the inhibitor sda |
| 6. | 12.4 | 3SL2.A | ATP forms a stable complex with the essential histidine kinase walk (yycg) domain |
| 7. | 13.6 | 1BXD.A | NMR structure of the histidine kinase domain of the e. coli osmosensor envz |
| 8. | 9.9 | 4EW8.A | Crystal structure of a c-terminal part of tyrosine kinase (divl) from *Caulobacter crescentus* cb15 at 2.50Å resolution (psi community target, shapiro) |
| 9. | 8e-14 | 4JAS.A | Structural basis of a rationally rewired protein-protein interface (HK853MUTANT A268V, A271G, T275M, V294T AND D297E AND RR468MUTANT V13P, L14I, I17M AND N21V) |
| 10. | 2e-10 | 4Q20.A | Crystal structure of a c-terminal part of tyrosine kinase (divl) from *Caulobacter crescentus* cb15 at 2.50Å resolution (psi community target, shapiro) |
| 11. | 2e-61 | 4PL9.A | Structure of the catalytic domain of ETR1 from *Arabidopsis thaliana* |

**Table 5. List of different templates selected by MOE based on Z scores from PDB, for ETR2 kinase domain protein sequence using search function in MOE and PDB BLAST.**

The table provides information on the full names of the templates along with their PDB ID and Z scores with ETR2 sequence. The chain name used as the template was indicated after the period in the PDB ID.

| SNo. | Z/E Score | PDB ID | Name of the Template |
|---|---|---|---|
| 1. | 6.8 | 3A0Y.B | Catalytic domain of histidine kinase thka (tm1359) (nucleotide free form 3: 1,2-propanediol, orthorombic) |
| 2. | 6.7 | 2C2A.A | Structure of the entire cytoplasmic portion of a sensor histidine kinase protein |
| 3. | 4e-10 | 4PL9.A | Structure of the catalytic domain of ETR1 from *Arabidopsis thaliana* |
| 4. | 30 | 3FEG.A | Crystal structure of human choline kinase beta in complex with phosphorylated hemicholinium-3 and adenosine nucleotide |
| 5. | 35 | 3LQ3.A | Crystal structure of human choline kinase beta in complex with phosphorylated hemicholinium-3 and adenosine nucleotide |
| 6. | 46 | 3WIQ.A | Crystal structure of kojibiose phosphorylase complexed with kojibiose |
| 7. | 67 | 4MYJ.A | Crystal Structure of PF3D7_1436600 |
| 8. | 77 | 4CMP.A | Crystal structure of *S. pyogenes* Cas9 |
| 9. | 77 | 4UN3.A | Crystal structure of Cas9 bound to PAM-containing DNA target |
| 10. | 77 | 4OO8.A | Crystal structure of Streptococcus pyogenes Cas9 in complex with guide RNA and target DNA |

**Table 6. List of different templates selected by MOE based on Z scores from PDB, for ERS2 kinase domain protein sequence using search function in MOE and PDB BLAST.**

The table provides information on the full names of the templates along with their PDB ID and Z scores with ERS2 sequence. The chain name used as the template was indicated after the period in the PDB ID.

| SNo. | Z/E Score | PDB ID | Name of the Template |
|---|---|---|---|
| 1. | 23 | 3D54.A | Stucture of PurLQS from *Thermotoga maritima* |
| 2. | 36 | 2IW1.A | Crystal structure of waag, a glycosyltransferase involved in lipopolysaccharide biosynthesis |
| 3. | 89 | 2IV7.A | Crystal structure of waag, a glycosyltransferase involved in lipopolysaccharide biosynthesis |
| 4. | 155 | 1RY7.A | Crystal Structure of the 3 Ig form of FGFR3c in complex with FGF1 |
| 5. | 186 | 3BGA.A | Crystal structure of beta-galactosidase from *Bacteroides thetaiotaomicron* VPI-5482 |
| 6. | 6e-06 | 4LP9.A | Structure of the catalytic domain of ETR1 from *Arabidopsis thaliana* |
| 7. | 200 | 2BW3.A | Three-dimensional structure of the hermes dna transposase |
| 8. | 219 | 4D1Q.A | Hermes transposase bound to its terminal inverted repeat |
| 9. | 224 | 3CVR.A | Crystal structure of the full length IpaH3 |
| 10. | 255 | 4JN5.A | Crystal structures of the first condensation domain of the CDA synthetase |

**Table 7. List of different templates selected by MOE based on Z scores from PDB, for EIN4 kinase domain protein sequence using search function in MOE and PDB BLAST.**

The table provides information on the full names of the templates along with their PDB ID and Z scores with EIN4 sequence. The chain name used as the template was indicated after the period in the PDB ID.

| SNo. | Z/E Score | PDB ID | Name of the Template |
|---|---|---|---|
| 1. | 8.7 | 2C2A.A | Structure of the entire cytoplasmic portion of a sensor histidine kinase protein |
| 2. | 6.9 | 4EW8.A | Crystal structure of a c-terminal part of tyrosine kinase (divl) from *Caulobacter crescentus* cb15 at 2.50Å resolution (psi community target, shapiro) |
| 3. | 17 | 3FIU.A | Nicotinamide mononucleotide synthetase is the key enzyme for an alternative route of NAD biosynthesis in *Francisella tularensis*. |
| 4. | 17 | 1OBH.A | Structural and mechanistic basis of pre- and posttransfer editing by leucyl-tRNA synthetase |
| 5. | 20 | 2V0C.A | Leucyl-trnasynthetase from *Thermos thermophilus* complexed with a sulphamoyl analogue of leucyl-adenylate in the synthetic site and an adduct of amp with 5-fluoro-1,3-dihydro-1-hydroxy-2,1-benzoxaborole (an2690) in the editing site |
| 6. | 63 | 4CIU.A | Crystal structure of *E. coli* ClpB |
| 7. | 65 | 3KD4.A | Crystal structure of a putative protease (bdi_1141) from *Parabacteroides distasonis* atcc 8503 at 2.00Å resolution |
| 8. | 106 | 1QVR.A | Crystal Structure Analysis of ClpB |
| 9. | 2e-08 | 4PL9.A | Structure of the catalytic domain of ETR1 from *Arabidopsis thaliana* |

**Table 8. List of different templates selected by MOE based on Z scores from PDB, for ETR1 GAF domain protein sequence using search function in MOE and PDB BLAST**.

The table provides information on the full names of the templates along with their PDB ID and Z scores with ETR1 sequence. The chain name used as the template was indicated after the period in the PDB ID.

| SNo. | Z/E Score | PDB ID | Name of the Template |
|---|---|---|---|
| 1. | 4.5 | 1QZY.A | Human Methionine Aminopeptidase in complex with bengamide inhibitor LAF153 and cobalt |
| 2. | 4.5 | 1B6A.A | Human methionine aminopeptidase 2 complexed with tnp-470 |
| 3. | 6.4 | 1EKB.B | The serine protease domain of enteropeptidase bound to inhibitor val-asp-asp-asp-asp-lys-chloromethane |
| 4. | 14 | 3V08.A | Crystal structure of Equine Serum Albumin |
| 5. | 32 | 1UI5.A | Crystal structure of gamma-butyrolactone receptor (ArpA like protein) |
| 6. | 35 | 2XVG.A | Crystal structure of alpha-xylosidase (gh31) from cellvibrio japonicus |
| 7. | 39 | 1UI6.A | Crystal structure of gamma-butyrolactone receptor (ArpA-like protein) |
| 8. | 40 | 4DNA.A | Crystal structure of putative glutathione reductase from *Sinorhizobium meliloti* 1021 |
| 9. | 61 | 1V0E.A | Endosialidase of bacteriophage k1f |

**Table 9. List of different templates selected by MOE based on Z scores from PDB, for ERS1 GAF domain protein sequence using search function in MOE and PDB BLAST.**

The table provides information on the full names of the templates along with their PDB ID and Z scores with ERS1 sequence. The chain name used as the template was indicated after the period in the PDB ID.

| SNo. | Z/E Score | PDB ID | Name of the Template |
|---|---|---|---|
| 1. | 9.4 | 3WNL.A | D308A mutant of *Bacillus circulans* T-3040 cyclo iso malto oligosaccharide glucanotransferase complexed with iso maltohexaose |
| 2. | 9.4 | 3WNK.A | Crystal Structure of *Bacillus circulans* T-3040 cyclo iso malto oligosaccharide glucanotransferase |
| 3. | 9.8 | 3WNP.A | D308A, F268V, D469Y, A513V, and Y515S quintuple mutant of *Bacillus circulans* T-3040 cyclo iso malto oligosaccharide glucanotransferase complexed with iso maltoundecaose |
| 4. | 16 | 3TTG.A | Crystal structure of putative amino methyltransferase from *Leptospirillum rubarum* |
| 5. | 35 | 4NQ1.A | *Legionella pneumophila* dihydrodipicolinate synthase with first substrate pyruvate bound in the active site |
| 6. | 54 | 2IUJ.A | Crystal Structure of Soybean Lipoxygenase-B |
| 7. | 60 | 4RH7.A | Crystal structure of human cytoplasmic dynein 2 motor domain in complex with ADP.Vi |
| 8. | 89 | 3FY4.A | (6-4) Photolyase Crystal Structure |
| 9. | 91 | 4KRE.A | Structure of Human Argonaute-1 bound to endogenous Sf9 RNA |
| 10. | 91 | 4KXT.A | Structure of human ARGONAUTE1 in complex with guide RNA |
| 11. | 127 | 4WJ3.M | Crystal structure of the asparagine transamidosome from Pseudomonas aeruginosa |
| 12. | 167 | 3V08.A | Crystal structure of Equine Serum Albumin |

**Table 10. List of different templates selected by MOE based on Z scores from PDB, for ETR2 GAF domain protein sequence using search function in MOE and PDB BLAST.**

The table provides information on the full names of the templates along with their PDB ID and Z scores with ETR2 sequence. The chain name used as the template was indicated after the period in the PDB ID.

| SNo. | Z/E Score | PDB ID | Name of the Template |
|---|---|---|---|
| 1. | Z – 6.3 | 4BWI.B | Structure of the phytochrome Cph2 from *Synechocystis* sp. PCC6803 |
| 2. | 7.7 | 2K2N.A | Solution structure of a cyanobacterial phytochrome GAF domain in the red light-absorbing ground state |
| 3. | 7.1 | 4IND.A | The Triple Jelly Roll Fold and Turret Assembly in an Archaeal Virus |
| 4. | E – 11 | 4LM1.A | Structure of the first RCC1-like domain of HERC2 |
| 5. | 23 | 3MPG.A | Solution structure of the [AibB8,LysB28,ProB29]-insulin analogue |
| 6. | 50 | 3Q9L.A | The structure of the dimeric E.coli MinD-ATP complex |
| 7. | 62 | 3WO1.A | Crystal structure of Trp332Ala mutant YwfE, an L-amino acid ligase, with bound ADP-Mg-Ala |
| 8. | 63 | 3WNZ.A | Crystal structure of Bacillus subtilis YwfE, an L-amino acid ligase, with bound ADP-Mg-Pi |
| 9. | 63 | 3A9U.A | Crystal structures and enzymatic mechanisms of a *Populus tomentosa* 4-coumarate--CoA ligase |
| 10. | 64 | 4HL7.A | Crystal structure of nicotinate phosphoribosyltransferase (target NYSGR-026035) from *Vibrio cholerae* |

**Table 11. List of different templates selected by MOE based on Z scores from PDB, for ERS2 GAF domain protein sequence using search function in MOE and PDB BLAST.**

The table provides information on the full names of the templates along with their PDB ID and Z scores with ERS2 sequence. The chain name used as the template was indicated after the period in the PDB ID.

| SNo. | Z/E Score | PDB ID | Name of the Template |
|---|---|---|---|
| 1. | 5.4 | 3ONM.A | Effector binding Domain of LysR-Type transcription factor RovM from *Y. pseudotuberculosis* |
| 2. | 7.6 | 3UCQ.A | Crystal structure of amylosucrase from *Deinococcus geothermalis* |
| 3. | 31 | 2LB5.A | Refined Structural Basis for the Photoconversion of A Phytochrome to the Activated FAR-RED LIGHT-ABSORBING Form |
| 4. | 34 | 1DG3.A | Structure of human guanylate binding protein-1 in nucleotide free form |
| 5. | 36 | 2K2N.A | Solution structure of a cyanobacterial phytochrome GAF domain in the red light-absorbing ground state |
| 6. | 42 | 1QY9.A | Crystal structure of *E. coli* Se-MET protein YDDE |
| 7. | 196 | 4MVF.A | Crystal Structure of *Plasmodium falciparum* CDPK2 complexed with inhibitor staurosporine |
| 8. | 202 | 1X1N.A | Structure determination and refinement at 1.8 A resolution of disproportionating enzyme from potato |
| 9. | 306 | 3RZG.A | Duplex interrogation by a direct DNA repair protein in the search of damage |
| 10. | 307 | 2K7W.A | BAX activation is Initiated at a novel interaction site |

**Table 12. List of different templates selected by MOE based on Z scores from PDB, for EIN4 GAF domain protein sequence using search function in MOE and PDB BLAST.**

The table provides information on the full names of the templates along with their PDB ID and Z scores with EIN4 sequence. The chain name used as the template was indicated after the period in the PDB ID.

| SNo. | Z/E Score | PDB ID | Name of the Template |
|---|---|---|---|
| 1. | 13<br>E | 3J61.Y | Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. |
| 2. | 24 | 3WRF.A | Structure and catalytic mechanism of a glycoside hydrolase Family-127 beta-L-arabinofuranosidase (HypBA1) |
| 3. | 27 | 1NM8.A | Structure of human carnitine acetyltransferase: molecular basis for fatty acyl transfer |
| 4. | 37 | 2K2N.A | Solution structure of a cyanobacterial phytochrome GAF domain in the red-light-absorbing ground state |
| 5. | 43 | 2LB5.A | Structural basis for the photoconversion of a phytochrome to the activated Pfr form. |
| 6. | 56 | 2PRR.A | Crystal structure of alkyl hydroperoxidaseAhpD core: uncharacterized peroxidase-related protein (YP_296737.1) from *Ralstonia eutropha* JMP134 at 2.15 A resolution |
| 7. | 75 | 3FCR.A | Crystal structure of putative aminotransferase (YP_614685.1) from *SILICIBACTER* SP. TM1040 at 1.80 A resolution |
| 8. | 96 | 3EFV.As | Crystal structure of a putative succinate-semialdehyde dehydrogenase from *Salmonella typhimurium* LT2 with bound NAD |
| 9. | 107 | 1YVU.A | Crystal structure of *A. aeolicus* argonaute |
| 10. | 130 | 4GQT.A | N-terminal domain of *C. elegans* Hsp90 |

# VITA

Sai Keerthana Wuppalapati was born in Nellore, Andhra Pradesh, India on June 5th, 1989, the only daughter of Sri. Murali Krishna Wuppalapati and Smt. Chengalamma Kaduru. After completing her work at Dr. Sarvepalli Radha Krishnan High School in Nellore, Andhra Pradesh she enrolled in the fall of 2005 in Sri Chaitanya Junior College for her higher school education at Vijayawada, Andhra Pradesh. She received her Bachelor of Science in Biomedical Sciences with a major in Human Genetics from Sri Ramachandra University, Chennai, India, where she was awarded "Founder Chancellor Cash Award" and "Mangalam Memorial Gold Medal" for being the best student of the batch 2007-2011 in the university. She then entered graduate school in fall of 2012 at The University of Tennessee at Knoxville.